MODULE 7 PART II

CONTINUOUS PROBABILITY MODELS

EXPECTED VALUE AND VARIANCE

SPECIAL PROBABILITY DENSITY FUNCTIONS

MODULE 7 PART II

CONTINUOUS PROBABILITY MODELS

CONTINUOUS PROBABILITY DISTRIBUTION

If the function f is a probability function with domain $\{x_1, x_2, \ldots, x_n\}$, and $f(x_i)$ is the probability that event x_i occurs, then for $1 \leq i \leq n$,

$$0 \leq f(x_i) \leq 1$$

and

$$f(x_1) + f(x_2) + \cdots + f(x_n) = 1$$

Discrete probability functions have a finite domain or an infinite domain that can be listed.

A continuous random variable can take on any value in some interval of real numbers. The distribution of this random variable is called a *continuous probability distribution*.

If X is a continuous random variable whose distribution is described by the function f on $[a, b]$, then

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

PROBABILITY DENSITY FUNCTION

The function f is a probability density function of a random variable X on the interval $[a, b]$ if

$f(x) \geq 0$ for all x in the interval $[a, b]$, and

$$\int_a^b f(x)\, dx = 1$$

Show that the function defined by $f(x) = (3/26)x^2$ is a probability density function for the interval $[1,3]$

Notice $f(x) \geq 0$ for all x in $[1,3]$

$$\int_1^3 \frac{3}{26} x^2\, dx = \frac{x^3}{26}\bigg|_1^3 = \frac{27-1}{26} = 1$$

Find $P(1 \leq X \leq 2)$

$$P(1 \leq X \leq 2) = \int_1^2 \frac{3}{26} x^2\, dx = \frac{x^3}{26}\bigg|_1^2 = \frac{7}{26}$$

EXAMPLE

Suppose the random variable X is the distance (in kilometers) from a given point to the nearest bird's nest, with the probability distribution function given by $f(x) = 2xe^{-x^2}$ for $x \geq 0$. Show that $f(x)$ is a probability density function.

Since $e^{-x^2} = 1/e^{x^2}$ is always positive, and $x \geq 0$, we have

$$f(x) = 2xe^{-x^2} \geq 0$$

Let $u = -x^2$, then $du = -2xdx$ and $-du = 2xdx$

$$\int 2xe^{-x^2}\,dx = -\int e^u\,du = -e^u = -e^{-x^2}$$

$$\int_0^\infty 2xe^{-x^2}\,dx = \lim_{b\to\infty}\int_0^b 2xe^{-x^2}\,dx = \lim_{b\to\infty}\left(-e^{-x^2}\right)\Big|_0^b$$

$$= \lim_{b\to\infty}\left(-\frac{1}{e^{b^2}} + e^0\right) = 0 + 1 = 1$$

Find the probability that there is a bird's nest within 0.5 km of the given point.

$$P(0 \leq X \leq 0.5) = \int_0^{0.5} 2xe^{-x^2}\,dx = \left(-e^{-x^2}\right)\Big|_0^{0.5}$$

$$= -e^{-(0.5)^2} - (-e^0) = -e^{-0.25} + 1$$

$$\approx -0.7788 + 1 = 0.2212$$

CUMULATIVE DISTRIBUTION FUNCTION

If f is a probability density function of a random variable in the interval $[a, b]$, then the cumulative distribution function is defined as

$$F(x) = P(X \leq x) = \int_a^x f(t)\, dt$$

for $x \geq a$. Also, $F(x) = 0$ for $x < a$.

Recall the probability density function in the previous example, $f(x) = 2xe^{-x^2}$ for $x \geq 0$

Find the cumulative distribution function

$$F(x) = P(X \leq x) = \int_0^x 2te^{-t^2}\, dt = -e^{-t^2} \Big|_0^x = -e^{-x^2} + 1$$

Use the CDF to calculate the probability that there is a bird's nest within 0.5 km of the given point.

To find $P(X \leq 0.5)$, calculate $F(0.5) = 1 - e^{-0.5^2} \approx 0.2212$

MODULE 7 PART II

EXPECTED VALUE AND VARIANCE

EXPECTED VALUE

Expected Value (Discrete Case)

Suppose the random variable X can take on the n values, $x_1, x_2, x_3, \dots x_n$. Also suppose the probabilities that each of these values occurs are $p_1, p_2, p_3, \dots p_n$, respectively. Then the mean, or expected value, of the random variable is

$$\mu = x_1p_1 + x_2p_2 + x_3p_3 + \cdots + x_np_n = \sum_{i=1}^{n} x_ip_i$$

Expected Value (Continuous Case)

If X is a continuous random variable with probability density function f on $[a, b]$, then the expected value of X is

$$E(X) = \mu = \int_a^b xf(x)\,dx$$

VARIANCE & STANDARD DEVIATION

If X is a continuous random variable with probability density function f on $[a, b]$, then the variance of X is

$$Var(X) = \sigma^2 = \int_a^b (x - \mu)^2 f(x)\, dx,$$

and the standard deviation of X is

$$\sigma = \sqrt{Var(X)}$$

<u>Note</u>: It is possible that $a = -\infty$ or $b = \infty$ in any of these formulas, in which case the density function f is defined on $[a, \infty), (-\infty, b]$, or $(-\infty, \infty)$

EXAMPLE

Find the expected value, variance, and standard deviation of the random variable X with probability density function $f(x) = (3/26)x^2$ on $[1,3]$.

Expected value

$$\mu = \int_1^3 xf(x)\,dx = \int_1^3 x\left(\frac{3}{26}x^2\right)dx$$

$$= \frac{3}{26}\int_1^3 x^3\,dx = \frac{3}{26}\left(\frac{x^4}{4}\right)\Big|_1^3$$

$$= \frac{3}{104}(81-1) = \frac{30}{13}$$

Variance

$$Var(X) = \int_1^3 \left(x - \frac{30}{13}\right)^2 \left(\frac{3}{26}x^2\right)dx$$

$$= \int_1^3 \left(x^2 - \frac{60}{13}x + \frac{900}{169}\right)\left(\frac{3}{26}x^2\right)dx$$

$$= \frac{3}{26}\int_1^3 \left(x^4 - \frac{60}{13}x^3 + \frac{900}{169}x^2\right)dx$$

$$= \frac{3}{26}\left(\frac{x^5}{5} - \frac{60}{13}\left(\frac{x^4}{4}\right) + \frac{900}{169}\left(\frac{x^3}{3}\right)\right)\Big|_1^3$$

$$= \frac{3}{26}\left[\left(\frac{243}{5} - \frac{60(81)}{52} + \frac{900(27)}{169(3)}\right) - \left(\frac{1}{5} - \frac{60}{52} + \frac{300}{169}\right)\right] \approx 0.2592$$

Standard deviation

$$\sigma \approx \sqrt{0.2592} \approx 0.5091$$

Alternative formula for variance:

$$Var(X) = \int_a^b x^2 f(x)\,dx - \mu^2$$

Note μ^2 is outside of the integrand

$$Var(X) = \int_1^3 x^2\left(\frac{3}{26}x^2\right)dx - \left(\frac{30}{13}\right)^2$$

Your turn: Find $Var(X)$ using the alternative formula.

APPLICATION

The clotting time of blood (in seconds) is a random variable with probability density function defined for t in $[1,20]$ by

$$f(t) = \frac{1}{(\ln 20)t}$$

Find the mean clotting time

$$\mu = \int_1^{20} t\left(\frac{1}{(\ln 20)t}\right) dt = \int_1^{20} \frac{1}{\ln 20}\, dt$$

$$= \frac{t}{\ln 20}\Bigg|_1^{20} = \frac{19}{\ln 20} \approx 6.342 \text{ sec}$$

Find the standard deviation

$$Var(T) = \int_1^{20} \frac{t^2}{(\ln 20)t}\, dt - \mu^2 = \int_1^{20} \frac{t}{\ln 20}\, dt - (6.342)^2$$

$$= \frac{1}{\ln 20}\left(\frac{t^2}{2}\right)\Bigg|_1^{20} - (6.342)^2 \approx 26.3687 \text{ sec}$$

$$\sigma = \sqrt{26.3687} \approx 5.135 \text{ sec}$$

Find the probability that a person's blood clotting time is within 1 standard deviation of the mean

$$\mu - \sigma = 6.342 - 5.135 = 1.2074$$
$$\mu + \sigma = 6.342 + 5.135 = 11.4774$$

$$P(1.2074 \le T \le 11.4774) = \int_{1.2074}^{11.4774} \frac{1}{(\ln 20)t}\, dt$$

$$= \frac{1}{\ln 20}(\ln 11.4774 - \ln 1.2074) \approx 0.7517$$

MODULE 7 PART II

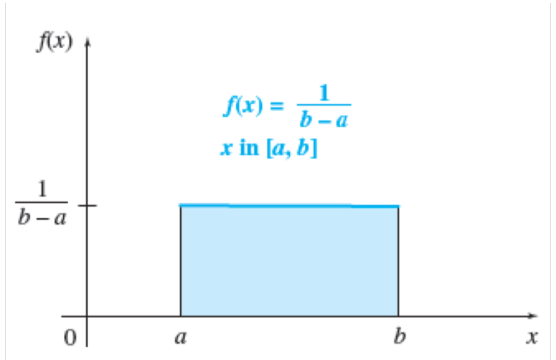
SPECIAL PROBABILITY DENSITY FUNCTIONS

UNIFORM DISTRIBUTION

The probability density function for the uniform distribution is defined by

$$f(x) = \frac{1}{b-a}$$

for x in $[a, b]$, where a and b are constant real numbers.



Uniform distribution

Notice $b - a > 0$, so $f(x) \geq 0$ and

$$\int_a^b \frac{1}{b-a}\,dx = \frac{1}{b-a}x\bigg|_a^b = \frac{1}{b-a}(b-a) = 1$$

$$\mu = \int_a^b \left(\frac{1}{b-a}\right) x\,dx = \left(\frac{1}{b-a}\right)\frac{x^2}{2}\bigg|_a^b$$

$$= \frac{1}{2(b-a)}(b^2 - a^2) = \frac{1}{2}(b+a)$$

$$Var(X) = \int_a^b \left(\frac{1}{b-a}\right) x^2\,dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{12}(b-a)^2$$

$$\sigma = \frac{1}{\sqrt{12}}(b-a)$$

<u>Your turn</u>: Verify the result for the variance

EXPONENTIAL DISTRIBUTION

If X is a random variable with probability density function $f(x) = ae^{-ax}$ for x in $[0, \infty)$, then $\mu = 1/a$ and $\sigma = 1/a$

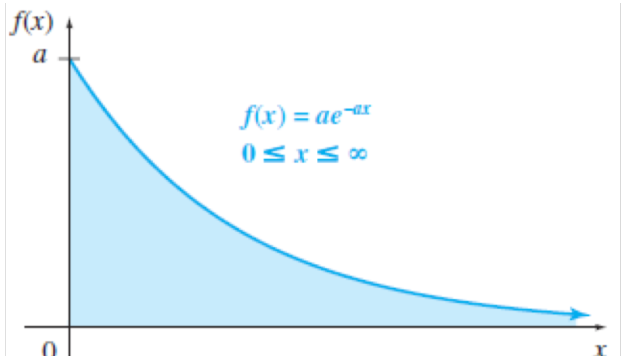
Suppose the useful life (in hours) of a flashlight battery is the random variable T, with probability density function given by

$$f(t) = \frac{1}{20}e^{-t/20}, \qquad t \geq 0$$

Find the probability that a particular battery, selected at random, has a useful life of less than 100 hours

$$P(T \leq 100) = \int_0^{100} \frac{1}{20}e^{-t/20}dt = \frac{1}{20}\left(-20e^{-t/20}\right)\Big|_0^{100}$$

$$= -(e^{-5} - e^0) \approx 1 - 0.0067 = 0.9933$$



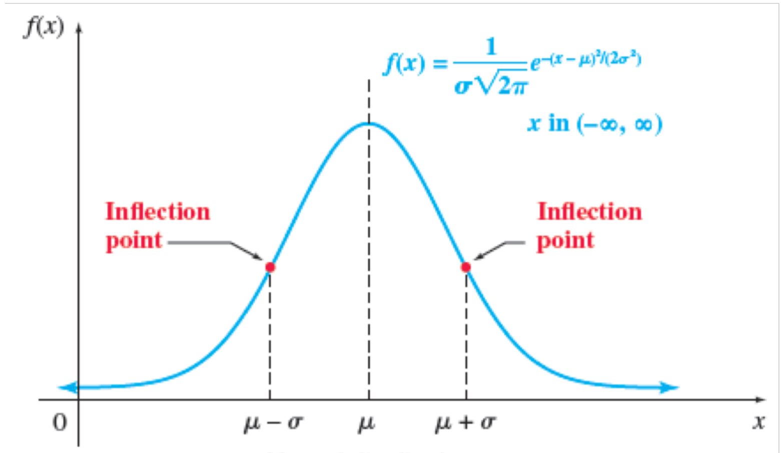
Exponential distribution

Find and interpret the mean and standard deviation

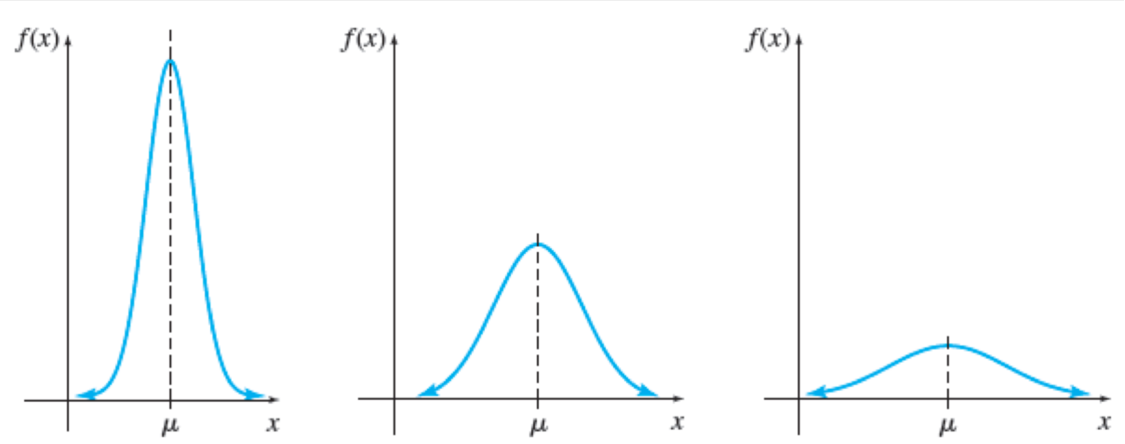
$\mu = 20$ and $\sigma = 20$

This means the average life of a battery is 20 hours and no battery lasts less than 1 standard deviation below the mean

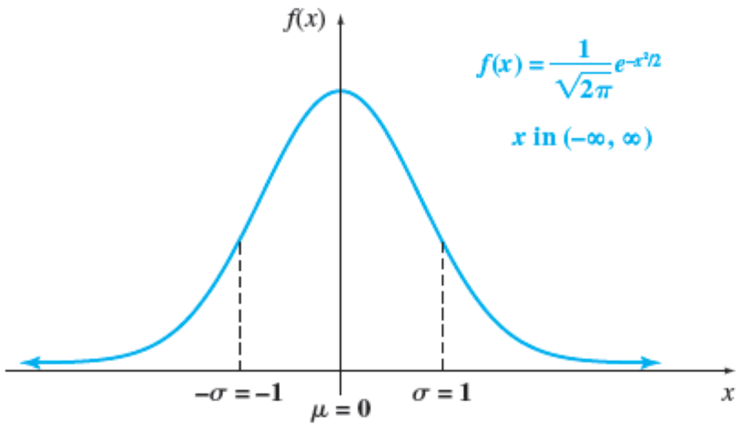
NORMAL DISTRIBUTION



Normal distribution



STANDARD NORMAL DISTRIBUTION



Standard normal distribution

<u>z-Scores Theorem</u>

Suppose a normal distribution has mean μ and standard deviation σ. The area under the associated normal curve that is to the left of the value x is equal to the area to the left of

$$z = \frac{x - \mu}{\sigma}$$

for the standard normal curve.

EXAMPLE

Life spans in a particular country are approximately normally distributed with a mean of about 75 years and a standard deviation of about 16 years.

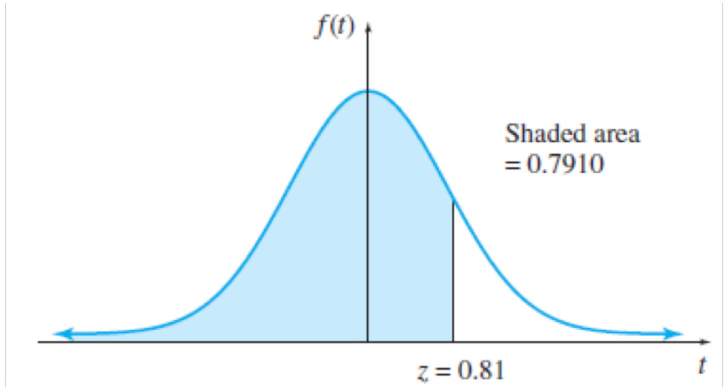
Find the probability that a randomly selected person from the country lives less than 88 years

$$z = \frac{88 - 75}{16} \approx 0.81$$

Refer to the table in Appendix H in Introductory Statistics

<u>Note</u>: The probabilities in the table represent the area under the curve from 0 to z

$$P(Z < 0.81) = 0.5 + 0.2910 = 0.7910$$

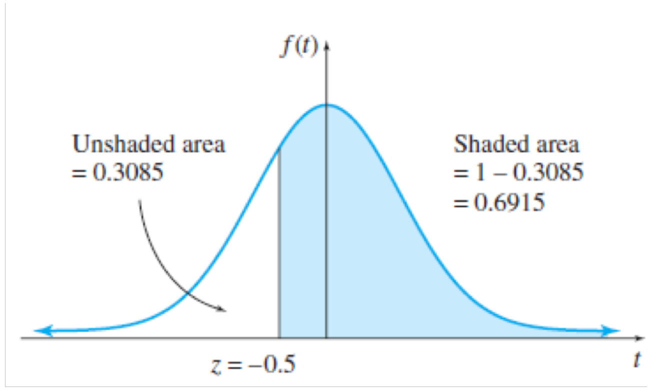


Find the probability that a randomly selected person lives more than 67 years

$$z = \frac{67 - 75}{16} \approx -0.5$$

Area to the left of $z = -0.5$ is $0.5 - 0.1915 = 0.3085$

$$P(Z > -0.5) = 1 - P(Z < -0.5)$$
$$= 1 - 0.3085 = 0.6915$$



EXAMPLE

Life spans in a particular country are approximately normally distributed with a mean of about 75 years and a standard deviation of about 16 years.

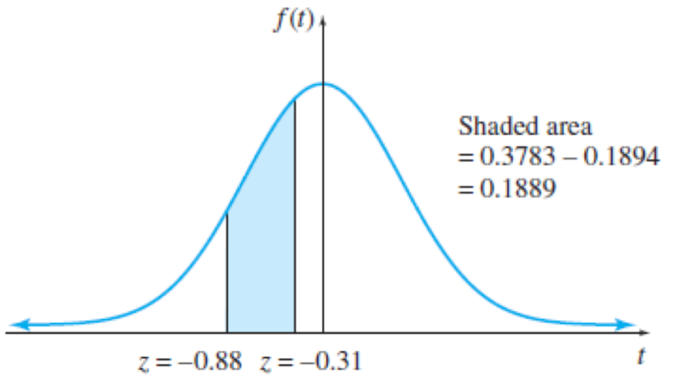
Find the probability that a randomly selected person lives between 61 and 70 years

$$z = \frac{61 - 75}{16} = -0.88 \quad \text{and} \quad z = \frac{70 - 75}{16} = -0.31$$

Area to the left of $z = -0.31$ is $0.5 - 0.1217 = 0.3783$

Area to the left of $z = -0.88$ is $0.5 - 0.3106 = 0.1894$

$P(-0.88 \leq Z \leq -0.31) = 0.3783 - 0.1894 = 0.1889$



Find a lower and upper life span that contain the middle 60% of the population

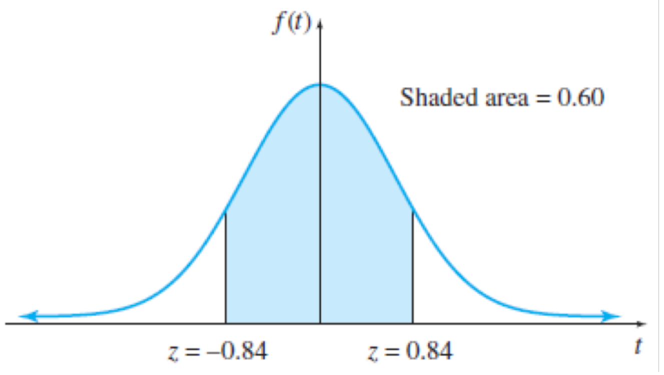
We need z-scores that capture the middle 60% of the distribution

Because of symmetry, this leaves 30% to the left of 0 and 30% to the right of 0

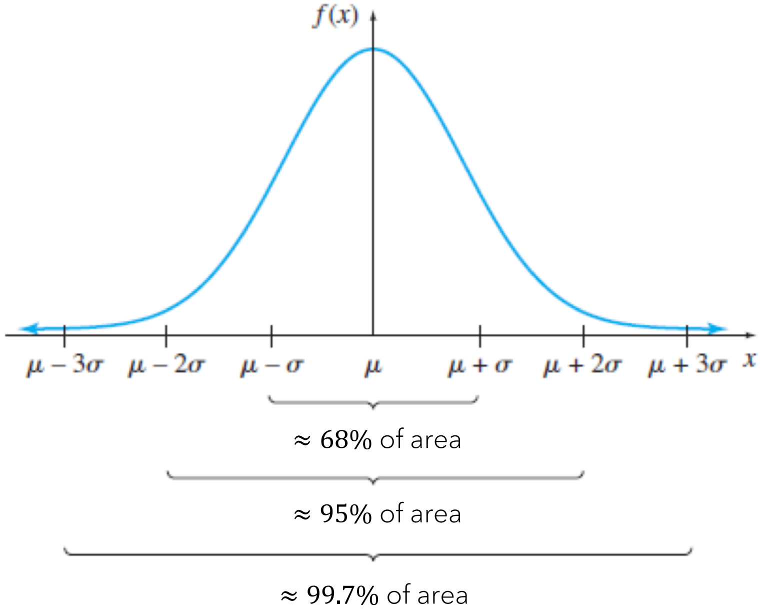
From the table, the best z-scores are ± 0.84

$z = (x - \mu)/\sigma$ so $x = \mu + z\sigma$

$x = 75 \pm (0.84)16$ so $x = 61.56$ and $x = 88.44$



EMPIRICAL RULE



Exam scores are normally distributed with a mean of 78 and standard deviation 6. Approximately what percent of scores lie between 66 and 90?

Approximately 68% of scores lie between 72 and 84

Approximately 95% of scores lie between 66 and 90

Approximately what percent of scores lie between 84 and 96?

$$\frac{1}{2}(0.997) = 0.4985$$

49.85% of scores lie between 78 and 96

$$\frac{1}{2}(0.68) = 0.34$$

34% of scores lie between 78 and 84

So, $49.85\% - 34\% = 15.85\%$ of scores lie between 84 and 96

QUESTIONS?