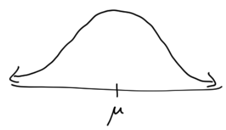
4.3 The Normal Distribution:

A continuous RV X has a normal distribution with parameters μ and σ^2 where $-\infty < \mu < \infty$ and $\sigma > 0$ if the pdf of X is

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



Shorthand: $X \sim N(\mu, \sigma^2)$

The Normal Distribution comes up very frequently because of the Central Limit Theorem which we will talk about in Ch. 5.

The idea is sums and averages of variables can often be approximated with a normal distribution.

Some Calculus will show that, if $X \sim N(\mu, \sigma^2)$,

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1, \quad E[X] = \mu, \text{ and } Var(X) = \sigma^2.$$

If $Z \sim N(0,1)$, then Z has the standard normal distribution.

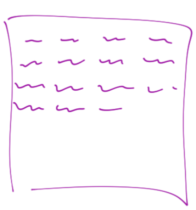
The pdf of a standard normal RV is

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad -\infty < y < \infty$$

Let $Z \sim N(0,1)$. What is $P(Z > 1)$?

$$P(Z > 1) = \int_1^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\,dy \approx 0.159$$

no way to express this integral in terms of nice functions.



table/computer

Since we can't explicitly calculate normal probabilities, we use technology or a table to approximate them.

Let $Z \sim N(0,1)$. The CDF of Z is often written Φ,

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy$$

Example 1: Let $Z \sim N(0,1)$. Use technology or a table to find

$P(Z \le 1.5)$, $P(-0.5 \le Z \le 0.5)$, and $P(Z \ge 1)$.

Φ is the CDF of Z

$$P(Z \le 1.5) = \Phi(1.5) \approx 0.9332$$

$$P(-0.5 \le Z \le 0.5) = \Phi(0.5) - \Phi(-0.5) \approx 0.383$$

$$P(Z \ge 1) = 1 - \Phi(1) \approx 0.159$$

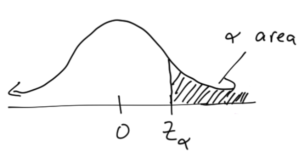
Example 2: Let $Z \sim N(0,1)$. What is the 90^{th} percentile of the standard normal distribution? In other words, what number z solves

$P(Z \le z) = 0.90$?

Using qnorm in R, we get $z \approx 1.28$

$$P(Z \le 1.28) \approx 0.90$$

Critical Values z_α: In statistics we will often want to use the value on the z axis for which α of the area under the curve lies to the right of z. $P(Z \ge z_\alpha) = \alpha$



Example 2 shows us that $z_{0.10} = 1.28$

Nonstandard Normal Distributions:

$$X = \sigma Z + \mu$$

When $X \sim N(\mu, \sigma^2)$, $\quad Z = \dfrac{X-\mu}{\sigma}$ has a standard normal distribution

Φ is the CDF of Z, $\Phi(y) = P(Z \le y)$

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

Example 3: The time it takes a driver to brake in response to brake lights is modeled by the RV $X \sim N(1.25, 0.5^2)$ with units in seconds.

What is $P(1 \le X \le 1.5)$? What is $P(X > 2)$?

For what value of x is $P(X \le x) = 0.95$?

One option is to use pnorm and qnorm in R, and just give R the mean and SD.

$P(1 \le X \le 1.5) =$ pnorm(1.5, mean = 1.25, sd = 0.5) − pnorm(1, mean = 1.25, sd = 0.5)

$P(X > 2) = 1 - P(X \le 2) = 1 -$ pnorm(2, mean = 1.25, sd = 0.5)

To find $P(X \le x) = 0.95$

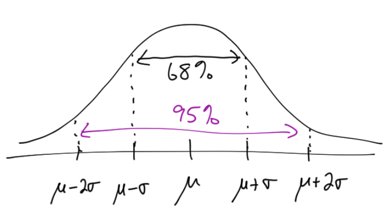
$x =$ qnorm(0.95, mean = 1.25, sd = 0.5)

The Empirical Rule: If the population distribution is approximately normal, then

Approximately 68% of the values are within 1 SD of the mean.

Approximately 95% of the values are within 2 SDs of the mean.

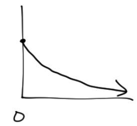
Approximately 99.7% of the values are within 3 SDs of the mean.



4.4 Exponential and Gamma Distributions:

X is said to have an exponential distribution with parameter $\lambda > 0$

if the pdf of X is $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$. $X \sim Exp(\lambda)$



Note: Some people use parameter $\beta = \frac{1}{\lambda}$ instead.

We can use integration by parts to find the mean and variance of

$X \sim Exp(\lambda)$.

$$E[X] = \int_0^\infty x \cdot \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$$

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

Let $x > 0$. Then the CDF is

$$F(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda y}\, dy = -e^{-\lambda y}\Big|_{y=0}^{y=x} = -e^{-\lambda x} + 1$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

Example 1: Say we model the time X between customers entering a store with an exponential distribution with parameter $\lambda = \frac{1}{5}$, with units in minutes.

Find $E[X]$, $P(X < 4)$, and $P(X \geq 5)$.

$$E[X] = \frac{1}{\lambda} = \frac{1}{\frac{1}{5}} = \boxed{5}$$

$$P(X < 4) = F(4) = \boxed{1 - e^{-\frac{1}{5} \cdot 4}} \approx \boxed{0.551}$$

$$P(X \geq 5) = 1 - F(5) = 1 - (1 - e^{-\frac{1}{5} \cdot 5}) = \boxed{e^{-1}} \approx \boxed{0.368}$$

$$P(X \le x)$$
$$F(x) = 1 - e^{-\lambda x}$$
$$1 - F(x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

Memoryless Property of the Exponential:

Let $X \sim \text{Exp}(\lambda)$. Let $t, t_0 > 0$. What is

$$P(X \ge t + t_0 \mid X \ge t_0) = \frac{P(X \ge t + t_0 \text{ and } X \ge t_0)}{P(X \ge t_0)}$$

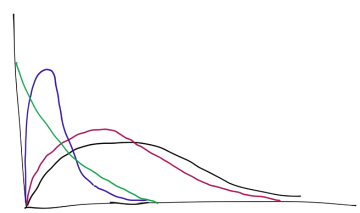
$$= \frac{P(X \ge t + t_0)}{P(X \ge t_0)} = \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t_0}} = \frac{e^{-\lambda t} e^{-\lambda t_0}}{e^{-\lambda t_0}} = e^{-\lambda t}$$

$$= P(X \ge t).$$

The Gamma Distribution:

A continuous RV X has the Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ if the pdf of X is

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$



where $\Gamma(\alpha)$ is the Gamma Function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx.$$

Shorthand: $X \sim \text{Gamma}(\alpha, \beta)$

Relationship with $\text{Exp}(\lambda)$ distribution:

$$\boxed{\text{pdf of } \text{Exp}(\lambda) \quad f(x) = \lambda e^{-\lambda x}, \quad x \ge 0}$$

$$\alpha = 1, \quad \beta = \frac{1}{\lambda}$$

$\text{Gamma}(1, \frac{1}{\lambda})$ is the same as $\text{Exp}(\lambda)$.

Relationship with Chi-Squared Distribution: Let ν ("nu") be a positive integer. The Gamma distribution with parameters $\alpha = \frac{\nu}{2}$ and $\beta = 2$ is called the Chi-Squared distribution with ν degrees of freedom.

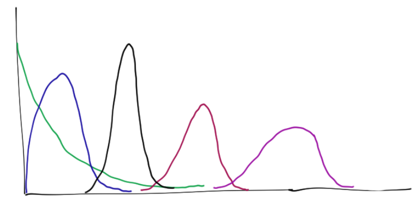
Shorthand: $X \sim \chi^2(\nu)$

4.5 Other Continuous Distributions

There are many other common continuous distributions. We'll focus on the Weibull distribution in this section. (The book also discusses the lognormal and beta distributions.)

A continuous RV X has the Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ if the pdf of X is

$$f(x) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$



Relationship with Exponential Distribution:

pdf for $Exp(\lambda)$ is
$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$

$\alpha = 1 \quad \beta = \frac{1}{\lambda}$

Weibull$\left(1, \frac{1}{\lambda}\right)$ is the same as $Exp(\lambda)$

$X \sim$ Weibull(α, β)

The CDF of the Weibull distribution can be found by integrating:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha}, & x \geq 0 \end{cases}$$

Example 1: Let X be the amount of nitric oxide emissions from a randomly selected four-stroke engine of a certain type. (Units g/gal)

Model $X \sim$ Weibull$(2, 10)$.

(a) Find $P(X \leq 10)$.

$$P(X \leq 10) = 1 - e^{-\left(\frac{10}{10}\right)^2} = \boxed{1 - e^{-1}} \approx \boxed{0.632}$$

(b) Find c so that $P(X > c) = 0.05$.

$$P(X \leq c) = 0.95$$
$$1 - e^{-\left(\frac{c}{10}\right)^2} = 0.95$$
$$e^{-\left(\frac{c}{10}\right)^2} = 0.05$$
$$-\left(\frac{c}{10}\right)^2 = \ln(0.05)$$
$$\left(\frac{c}{10}\right)^2 = -\ln(0.05)$$
$$\frac{c}{10} = \sqrt{-\ln(0.05)}$$
$$\boxed{c = 10\sqrt{-\ln(0.05)}} \approx \boxed{17.31}$$