Normal Distributions, chapter 5.6

Grethe Hystad

November 5, 2013

The Normal distribution

- The normal distribution is the most frequently occuring continuous probability distributions.
- Much statistics theory is based on the normal distribution.
- Central limit theorem

Examples:

- Heights of adults
- Observations errors in an experiment
- Velocities of the molecules in the ideal gas.
- Many naturally occuring measurements have relative frequency distributions that closely fit the normal curve.

The Normal distribution

Theorem

The Normal distribution:

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

$$E(X) = \mu \;\; \textit{and} \;\; V(X) = \sigma^2.$$

The Normal distribution

Theorem

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}.$$

Theorem

If X has a normal distribution with mean μ and a variance of σ^2, and

$$Y = aX + b,$$

where a, b are constants, then Y also has a normal distribution with

$$E(Y) = a\mu + b \;\; \text{and} \;\; V(Y) = a^2\sigma^2.$$

The Standard Normal distribution

Definition

We say that Z has a standard normal distribution if it has a normal distribution with $\mu = 0$ and $\sigma = 1$.

Define

$$Z = \frac{X - \mu}{\sigma},$$

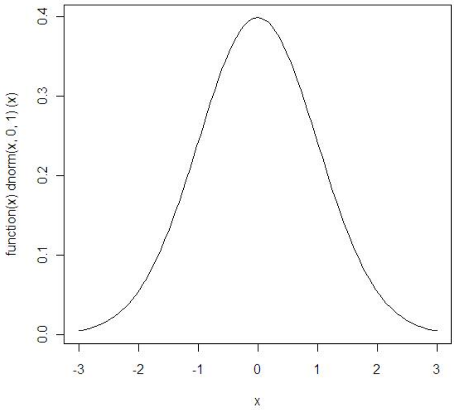
where X is a random variable that have a normal distribution with mean μ and standard deviation σ. Then Z has a standard normal distribution.

Tables of integrals, back in the book, gives numerically values for

$$P(0 \leq Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}}\, dt.$$

The Standard Normal distribution

```
> plot(function(x) dnorm(x,0,1),-3,3)
```



The Normal distribution

- The probability density function, $f(x)$, of a normal random variable with a mean of μ and a standard deviation of σ has points of inflection at $x = \mu \pm \sigma$. The proof of this is left as a homework problem.

Example

Let Z denote a standard normal variable. Find
(A) $P(Z \leq 1)$
(B) $P(Z \geq 1)$
(C) $P(Z < -1.5)$
(D) $P(-1 \leq Z \leq 1.5)$
(E) Find a value z_0 such that $P(Z \leq z_0) = 0.85$.

The Normal distribution

Solution

(A)

$$P(Z \leq 1) = P(Z \leq 0) + P(0 \leq Z \leq 1) = 0.5 + 0.3413 = 0.8413.$$

In R:

$>$ *pnorm(1)*

[1] 0.8413447

(B)

$$P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587.$$

In R:

$>$ *pnorm(1,lower.tail=FALSE)*

[1] 0.1586553

The Normal distribution, Solution cont.

Solution

(C)

$$\begin{aligned} P(Z < -1.5) &= P(Z > 1.5) \\ &= 1 - (0.5 + P(0 \leq Z \leq 1.5)) \\ &= 0.5 - 0.4332 = 0.0668. \end{aligned}$$

In R:

> pnorm(-1.5)

[1] 0.0668072

The Normal distribution, Solution cont.

Solution

(D)

$$\begin{aligned} P(-1 \leq Z \leq 1.5) &= P(-1 \leq Z \leq 0) + P(0 \leq Z \leq 1.5) \\ &= P(0 \leq Z \leq 1) + P(0 \leq Z \leq 1.5) \\ &= 0.3413 + 0.4332 = 0.7745. \end{aligned}$$

In R:

```
> pnorm(1.5)-pnorm(-1)
[1] 0.7745375
```

The Normal distribution, Solution cont.

Solution

(E) By table:

$$P(Z \leq z_0) = 0.5 + P(0 \leq Z \leq z_0) = 0.85.$$

Hence

$$P(0 \leq Z \leq z_0) = 0.35.$$

From the table, we find $z_0 = 1.04$*.*
In R:
$> z_0 = qnorm(0.85)$
[1] 1.0364.

In R

- The function **qnorm**(p, μ, σ) returns the *quantile* for which there is a probability of p of getting a value less than or equal to it. Thus, the *quantile* is the value x such that $P(X \leq x) = p$ for a given p.
- The function **pnorm**(x, μ, σ) computes the proportion of observations in the normal distribution that are less than or equal to x; that is $P(X \leq x)$, where X is $N(\mu, \sigma)$. I other words, **qnorm** converts proportions to quantiles while **pnorm** converts quantiles to proportions which means that **qnorm** and **pnorm** are inverse functions of each other.
- The function **pnorm**$(x, \mu, \sigma, lower.tail = FALSE)$ computes the proportion of observations in the normal distribution that are greater than or equal to x; that is $P(X \geq x)$, where X is $N(\mu, \sigma)$.

In R

- The density for a continuous distribution measures the probability of getting a value close to x. Continuous random variables have a density at a point since they have no probability at a single point, $P(X = x)$. For the normal distribution we compute the density using the function **dnorm**(x, μ, σ).

The Normal distribution

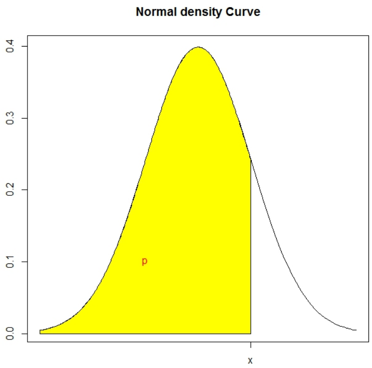


Figure : Normal density curve illustrating $P(X \leq x) = p$

The Standard Normal distribution, the R-code

If $P(Z \leq z_0) = p$, then in R,
$> z_0 = qnorm(p)$.
To compute $F(z_0) = \int_{-\infty}^{z_0} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$, in R:
$> pnorm(z_0)$.

The Standard Normal distribution

We have

$$P(-1 \leq Z \leq 1) = 2P(0 \leq Z \leq 1) = 2 \cdot 0.3413 = 0.683.$$

$$P(-2 \leq Z \leq 2) = 2 \cdot 0.4772 = 0.954.$$

$$P(-3 \leq Z \leq 3) = 2 \cdot 0.4987 = 0.997.$$

Thus, for a standard normal distribution, approximately,
68% of the values fall within 1 standard deviation of the mean in either direction.
95% of the values fall within 2 standard deviation of the mean in either direction.
99.7% of the values fall within 3 standard deviation of the mean in either direction.

The Normal distribution

Example

- Find $P(4 \leq X \leq 12)$, where X is a normal random variable that has a mean of 8 and a standard deviation of 2.

The Normal distribution

Example

- Find $P(4 \leq X \leq 12)$, where X is a normal random variable that has a mean of 8 and a standard deviation of 2.

Solution

Let

$$Z = \frac{X-8}{2}.$$

Then

$$P(4 \leq X \leq 12) = P\left(\frac{4-8}{2} \leq Z \leq \frac{12-8}{2}\right)$$
$$= P(-2 \leq Z \leq 2) = 0.954.$$

The Normal distribution, R-code

In R, the problem can also be solved in this way:

$$P(4 \leq X \leq 12) = P(X \leq 12) - P(X \leq 4).$$

Then type
$>$ pnorm(12,8,2)-pnorm(4,8,2)
[1] 0.954.
In general, in R:
$> pnorm(x, \mu, \sigma)$
which is the cumulative normal distribution, $P(X \leq x)$, with mean μ and standard deviation σ.

Problem 3.35 from The Basic Practice of Statistics by Moore et al., 2013

In its Fuel Economy Guide for model year 2010 vehicles, the Environmental Protection Agency gives data on 1101 vehicles. There are a number of high outliers, mainly hybrid gas-electric vehicles. If we ignore the vehicles identified as outliers, however, the combined city and highway gas mileage of the other 1082 vehicles is approximately Normal with mean 20.3 miles per gallon (mpg) and standard deviation 4.3 mpg.
The 2010 Chevrolet Camaro with an eight-cylinder engine and automatic transmission has a combined gas mileage of 19 mpg. What percent of all vehicles have better gas mileage than the Camaro?

Example

If X has a normal distribution with mean 2 and standard deviation 2, find $P(X^2 - 4X \leq 12)$.

Example

If X has a normal distribution with mean 2 and standard deviation 2, find $P(X^2 - 4X \leq 12)$.

Solution

$$\begin{aligned} P(X^2 - 4X \leq 12) &= P(X^2 - 4X + 4 \leq 16) \\ &= P((X-2)^2 \leq 16) \\ &= P(-4 \leq (X-2) \leq 4) \\ &= P(-2 \leq \frac{X-2}{2} \leq 2) \\ &= P(-2 \leq Z \leq 2) = 0.95, \end{aligned}$$

where $Z = \frac{X-2}{2}$.