Continuous Uniform and Normal Distributions

Anastasiia Kim

February 26/28, 2020

Continuous Uniform

A continuous Uniform random variable X has a probability density function

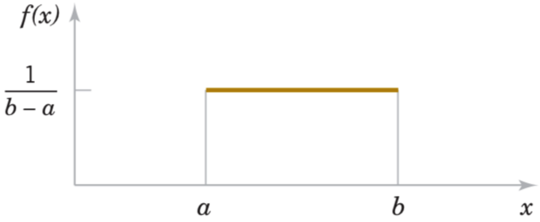
$$f(x) = \frac{1}{b-a}, \quad a \le x \le b$$

with $E(X) = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$.
The cumulative distribution function of a continuous uniform random variable is obtained by integration. If $a \le x < b$,

$$F(x) = \int_a^x \frac{1}{b-a} dy = \frac{x-a}{b-a}$$

$F(x) = 0, x < a$, and $F(x) = 1, x \ge b$

Density Curves



A density curve is a curve that:

- is always on or above the horizontal axis
- has an area of exactly 1 underneath it
- A density curve describes the overall pattern of a distribution.
- The area under the curve and above any range of values on the horizontal axis is the proportion of all observations that fall in that range.

Example

A beacon transmits a signal every 10 minutes (such as 8:20, 8:30, etc.). The time at which a receiver is tuned to detect the beacon is a continuous uniform distribution from 8:00 a.m. to 9:00 a.m. Consider the waiting time until the next signal from the beacon is received (X).

- Waiting time is always between 0 and 10 minutes, and it is uniformly distributed, so $X \sim Uniform(0, 10)$
- What is the mean waiting time?

$$E(X) = \frac{0 + 10}{2} = 5$$

- What is the probability that the waiting time is less than 3 minutes?

$$P(X < 3) = F(3) = \frac{3 - 0}{10 - 0} = 0.3$$

Example

An e-mail message will arrive at a time uniformly distributed between 9:00 a.m. and 11:00 a.m. You check e-mail at 9:15 a.m. and every 30 minutes afterward.

- What is the standard deviation of arrival time (in minutes)?
- What is the probability that the message arrives less than 10 minutes before you view it?

Normal (Gaussian) Distribution

- One particularly important class of density curves are the Normal curves, which describe Normal distributions.
- All Normal curves are symmetric, single-peaked, and bell-shaped.
- A Specific Normal curve is described by its mean μ and standard deviation σ.

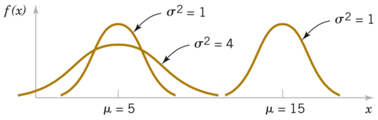


Figure 1: Normal probability density curves

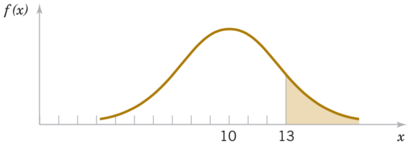


Figure 2: Probability that $X > 13$ for a normal r.v. with $\mu = 10, \sigma^2 = 4$

Normal (Gaussian) Distribution

A random variable X with probability density function

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

is a normal random variable with parameters $-\infty < \mu < \infty$ and $\sigma > 0$ with $E(X) = \mu$ and $Var(X) = \sigma^2$.
The notation $N(\mu, \sigma^2)$ is used to denote the distribution.

The 68-95-99.7 Rule

In the Normal distribution with mean μ and standard deviation σ:

- Approximately 68% of the observations fall within σ of μ.
- Approximately 95% of the observations fall within 2σ of μ.
- Approximately 99.7% of the observations fall within 3σ of μ.

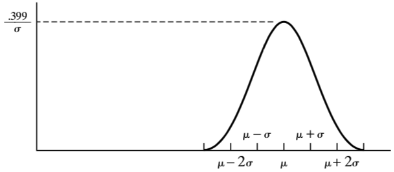


Figure 3: Normal density curve for μ, σ^2

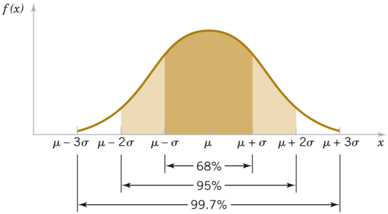


Figure 4: Illustration of the 68-95-99.7 rule

Normal Density Curve. Example

The distribution of Iowa Test of Basic Skills (ITBS) vocabulary scores for 7th-grade students in Gary, Indiana, is close to Normal. Suppose the distribution is $N(6.84, 1.55)$.

- Sketch the Normal density curve for this distribution.
- What percent of ITBS vocabulary scores are less than 3.74?
- What percent of the scores are between 5.29 and 9.94?

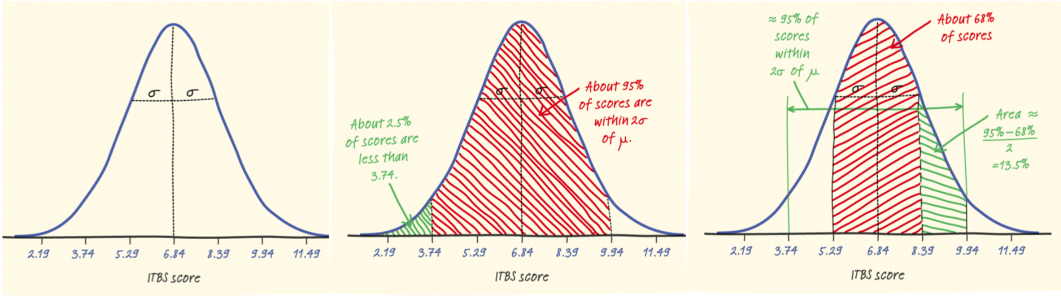


Figure 5: from The Basic Practice of Statistics (7th Edition), by Moore, Notz and Fligner

Standard Normal (Gaussian) Distribution

If $X \sim N(\mu, \sigma^2)$ then

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

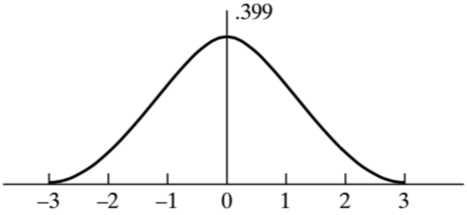
is normally distributed with parameters $\mu = 0$ and $\sigma^2 = 1$ and pdf for Z is

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

The cumulative distribution function of a standard normal r. v. is denoted as

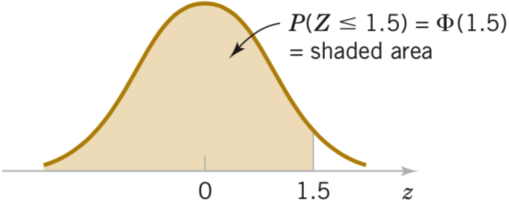
$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

$$\Phi(-z) = 1 - \Phi(z), \quad -\infty < z < \infty$$



Standard Normal (Gaussian) Distribution

Assume that Z is a standard normal random variable. Standard normal table or *pnorm*() in R can calculate $\Phi(z) = P(Z \leq z)$



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

- Read down the z column to the row that equals 1.5. The probability is read from the adjacent column, labeled 0.00, to be 0.93319. Also using R: $pnorm(1.5) = 0.9331928$ or $pnorm(1.5, 0, 1) = 0.9331928$
- The column headings refer to the hundredths digit of the value of z in $P(Z \leq z)$. For example, $P(Z \leq 1.53)$ is found by reading down the z column to the row 1.5 and then selecting the probability from the column labeled 0.03 to be 0.93699.

Example

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

- $P(X > 0) = P(\frac{X-\mu}{\sigma} > \frac{0-\mu}{\sigma}) = P(Z > \frac{0-3}{3}) = P(Z > -1) = 1 - P(Z \leq -1) = 1 - \Phi(-1) = 1 - 0.159 = 0.841$. Since $pnorm(-1) = 0.159$
- $P(2 < X < 5)$
- $P(|X - 3| > 6)$

Example

The time until recharge for a battery in a laptop computer under common conditions is normally distributed with a mean of 260 minutes and a standard deviation of 50 minutes.

- What is the probability that a battery lasts more than four hours?
- What are the quartiles (the 25% (i.e. $P(X < x_1) = 0.25$) and 75% values) of battery life?
- What value of life in minutes is exceeded with 95% probability?

Normal approximation to the Binomial distribution

DeMoivre-Laplace limit theorem: when n is large, a binomial random variable with parameters n and p will have approximately the same distribution as a normal random variable with the same mean and variance as the binomial.
If X is a binomial random variable with parameters n and p: $X \sim Binomial(n, p)$,

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard normal random variable.
The normal approximation will, in general, be quite good for values of n satisfying $np(1-p) \geq 10$.
Note: we now have two possible approximations to binomial probabilities: the Poisson approximation, which is good when n is large and p is small, and the normal approximation, which can be shown to be quite good when $np(1-p)$ is large.

Normal approximation to the Binomial distribution

Let X be the number of times that a fair coin that is flipped 40 times lands on heads. Find the probability that X = 20. Use the normal approximation and then compare it with the exact solution.

$X \sim Binomial(40, 0.5)$ and $P(X = 20) = \binom{40}{20}(0.5)^40(0.5)^0 = 0.1254$

Before using the Normal approximation we need to make the continuity correction

$$P(X = x) = P(x - 0.5 < X < x + 0.5)$$

Therefore,

$$P(X = 20) = P(19.5 \leq X \leq 20.5) = P\left(\frac{19.5 - 20}{\sqrt{20(1 - 0.5)}} \leq \frac{X - np}{\sqrt{np(1 - p)}} \leq \frac{20.5 - 20}{\sqrt{20(1 - 0.5)}}\right) =$$

$$= \Phi(0.16) - \Phi(-0.16) = 0.1272$$

Example

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? Assume that if the diet has no effect on the cholesterol count, then, strictly by chance, each person's count will be lower than it was before the diet with probability 0.5.

Normal Approximation to the Poisson Distribution

If X is a Poisson random variable with rate λ ($E(X) = \lambda$, $Var(X) = \lambda$):
$X \sim Poisson(\lambda)$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard normal random variable.
The same continuity correction used for the binomial distribution can also be applied.
The approximation is good for $\lambda > 5$.

Example

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?
This probability can be expressed exactly as

$$P(X \leq 950) = \sum_{x=0}^{950} \frac{e^{-1000}(1000)^x}{x!}$$

The probability can be approximated as

$$P(X \leq 950) = P(X \leq 950.50) = P(Z \leq \frac{950.5 - 1000}{\sqrt{1000}}) = P(Z \leq -1.57) = 0.058$$

Example

A high-volume printer produces minor print-quality errors on a test pattern of 1000 pages of text according to a Poisson distribution with a mean of 0.4 per page.

- Why are the numbers of errors on each page independent random variables?
- What is the mean number of pages with errors (one or more)?
- Approximate the probability that more than 350 pages contain errors (one or more).