Statistics 301: Probability and Statistics

Continuous Distributions

Module 5

Updated 2019

Review of discrete random distributions

From Module 4, the distributions were discrete. A functions associated with a discrete random variable is usually called a *probability mass function (pmf).* The name pmf is suggested by a model used in physics for a system of "point masses". The pmf describes how the total probability mass of 1 is distributed at various points along the axis of possible values of the random variable.

Continuous distributions

Continuous Random Variable

- A random variable X is called **continuous** if it can take any value within a finite or infinite interval of the real number line $(-\infty, \infty)$
- Some examples would be measurements of length, strength, lifetime, pH, etc.

Density Function of a Continuous Random Variable

A continuous random variable X cannot have a PMF (probability mass function). The reason for this is that:

$P(X = x) = 0$

That is, there is no area under a curve at a single point

Probability Density Function (pdf)

Probability Density Function (pdf):

The probability density function (pdf) of a continuous random variable X is a nonnegative function f_X with the property that $P(a < X < b)$ equals the area under it and above the interval $[a, b]$. Thus,

$$P(a < X < b) = \text{area under } f_X$$

$$a \leq X \leq b$$

Generic pdf

The area under a curve is found by integration, such as:

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

Keep in mind that the above is equal to the following as you cannot find probabilities of a single point:

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

Rules of Probibility

1. $0 \le p(x) \le 1$
2. $\int_{-\infty}^{\infty} f(x)\ dx = 1$
3. Complement Rule
4. Addition Rule (for disjoint and non-disjoint events)
5. Multiplication Rule (for independent events)
6. Conditional Probability Rule

All the previously learned rules apply to continuous distributions.

Cumulative Distribution Function

The cumulative distribution function, or **CDF**, of a random variable X gives the probability of events of the form $[X \le x]$, for all numbers x

Notation for the cumulative distribution function is:

CDF or $F_X(x) = P(X \le x)$

$$F_X(x) = \int_{-\infty}^{x} f(y)dy$$

EX, VX, SDX of Generic pdf

- Expected Value (mean), variance, and standard deviation

$$EX = \int_{-\infty}^{\infty} xf(x)dx$$

$$VX = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

$$= E(X^2) - (EX)^2 \text{ where } E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$

$$SDX = \sqrt{VX}$$

Generic Example I

Let X denote the resistance of a randomly chosen resistor and suppose its pdf is given as:

$$f(x) = \begin{cases} kx & 8 \le x \le 10 \\ 0 & otherwise \end{cases}$$

(1) Calculate k
(2) CDF of X
(3) Calculate $P(X < 9)$
(4) Use CDF to calculate $P(8.6 \le X \le 9.8)$ and $P(X \le 9.8 | X \ge 8.6)$
(5) Calculate EX, VX, SDX

Generic Example II

$$\int_{8}^{10} kx\ dx$$

$$= \frac{1}{\frac{1}{2}(10^2 - 8^2)} \Rightarrow k = \frac{1}{18}$$

$$f(x) = \begin{cases} \frac{1}{8}x & 8 \leq x \leq 10 \\ 0 & otherwise \end{cases}$$

The CDF:

$F_X(x) = \int_8^x \frac{1}{18}y\ dy \Rightarrow \frac{1}{18}(\frac{y^2}{2})|_8^x$

$\Rightarrow \frac{1}{36}(x^2 - 64) = \frac{x^2-64}{36}$

Generic Example III

- $P(X < 9)$

$= \int_8^9 f(x)dx = \int_8^9 \frac{x}{8}dx = F(9) = \frac{9^2-64}{36} = 0.472222$

Generic Example IV

- $P(8.6 \leq X \leq 9.8)$

$= \int_{8.6}^{9.8} f(x)dx = F(9.8) - F(8.6) = \frac{9.8^2-64}{36} - \frac{8.6^2-64}{36} =$

$= 0.613333$

- For $P(X \leq 9.8|X \geq 8.6)$, formula for conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ so

$P(X \leq 9.8|X \geq 8.6) = \frac{P(X \leq 9.8 \cap X \geq 8.6)}{P(X \geq 8.6)}$

$= \frac{P(8.6 \leq X \leq 9.8)}{P(X \geq 8.6)}$

$= \frac{F(9.8)-F(8.6)}{F(10)-F(8.6)} = 0.8479$

Generic Example V

- Find EX, VX, SDX

$$EX = \int_8^{10} x\left(\frac{1}{18}x\right)dx = \frac{1}{18}\int_8^{10} x^2 dx$$

$= \frac{1}{18}(\frac{x^3}{3})_8^{10} = \frac{10^3-8^3}{54} = 9.037037$

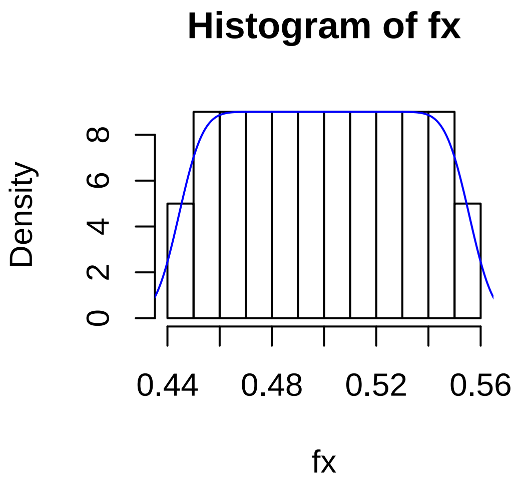
$$E(X^2) = \int_8^{10} x^2\left(\frac{1}{18}x\right)dx = 82$$

$VX = E(X^2) - (EX)^2 = 0.331962$

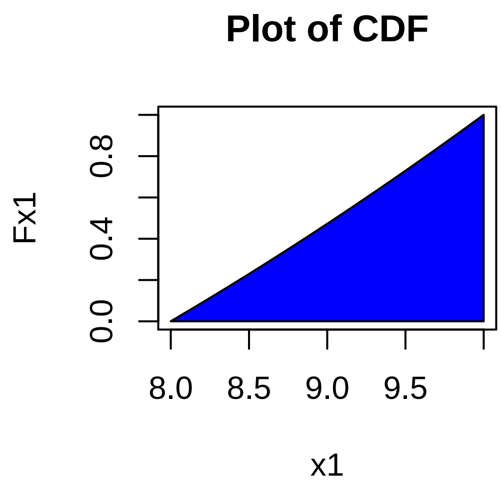
$SDX = \sqrt{VX} = 0.576161$

Graph of Generic Example pdf

Since there are an infinite number of values within the interval from 8 to 10, I just chose 1000 values (it could have been smaller but I just chose 1000)



Graph of Generic Example CDF



Uniform Distribution

In the context of probability distributions, a uniform distribution refers to a probability distribution for which all of the values that a random variable can take on occur with equal probability. This probability distribution is defined as follows.

A random variable X is said to be uniform if all values of X are equally likely

$$X \sim U(A, B)$$

$$P(a < X < b) = \int_a^b \frac{1}{B - A} dx \quad for \quad A < X < B$$

Uniform EX, VX, SDX

$$EX = \frac{B + A}{2}$$
$$VX = \frac{(B - A)^2}{12}$$
$$SDX = \sqrt{VX}$$

Uniform Example I

Say that Y has a uniform distribution on the interval $[2, 5]$. Find the following:

- $f(y)$
- $F_Y(y)$
- $P(2 < Y < 3)$
- EX, VX, SDX

Uniform Example II

$$f(y) = \begin{cases} \frac{1}{5-2} & 2 \le y \le 5 \\ 0 & otherwise \end{cases}$$

$$F_Y(y) = \int_2^y \frac{1}{3} dx = \frac{1}{3}(x)_2^y =$$

$\frac{y-2}{3}$

$$P(2 < Y < 3)$$

$= F(3) - F(2) = 0.333333$

Uniform Example III

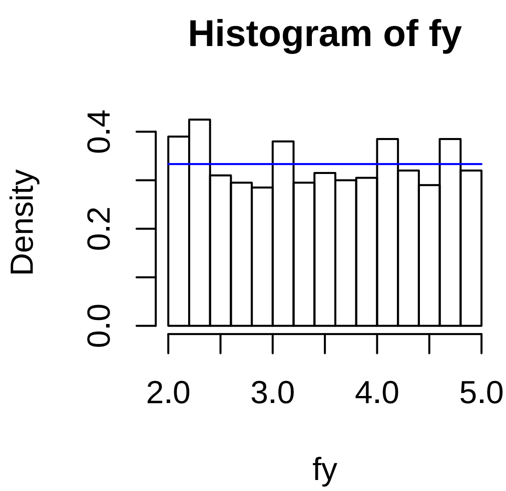
$$EX = \frac{B + A}{2} = \frac{5 + 2}{2} = 3.5$$

$$VX = \frac{(B - A)^2}{12} = \frac{(5 - 2)^2}{12} = 0.75$$

$$SDX = \sqrt{VX} = \sqrt{0.75} = 0.866025$$

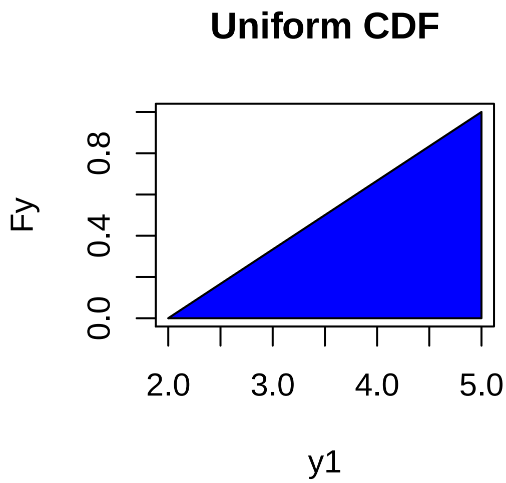
Graph of Uniform pdf

```
y1=seq(from=2,to=5,length.out=1000); fy=runif(length(y1),min=2,max=5)
hist(fy,prob=T); curve(dunif(x,min=2,max=5),col='blue',add=T)
```



Graph of Uniform CDF

```
y1=seq(from=2,to=5,length.out=1000); F=antiD((1/3)~y); Fy=F(y1)-F(2)
plot(y1,Fy,type='l',main="Uniform CDF")
polygon(c(y1,y1[length(y1)]), c(Fy,Fy[1]),col='blue')
```



The Exponential Distribution

X is said to have an **exponential distribution** with parameter λ and pdf:

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

and will always have the CDF (through integration by parts):

$$F_X(x) = 1 - e^{-\lambda x}$$

$$EX = \frac{1}{\lambda}$$
$$VX = \frac{1}{\lambda^2}$$
$$SDX = \sqrt{VX}$$

Exponential Example I

Suppose that the useful time (in years) of a PC is exponentially distributed with parameter $\lambda = 0.25$. A student entering a four-year undergraduate program inherits a two-year old PC from his sister who just graduated. Find the probability that the useful lifetime of the PC will last at least until he graduates (assume within 4 years). Let X denote the useful lifetime of the PC.

$f(x;\lambda) = f(x;0.25) = \frac{1}{4}e^{-\frac{1}{4}x}$

$P(X > 4+2|X > 2) = \frac{P(X>4+2 \cap X>2)}{P(X>2)}$

$= \frac{P(X>6)}{P(X>2)} = \frac{e^{-.25*6}}{e^{-.25*2}} = 0.367879$

Exponential Example II

About how long would you expect the PC to last, on average? (this is the question to find the mean). Find EX, VX, SDX

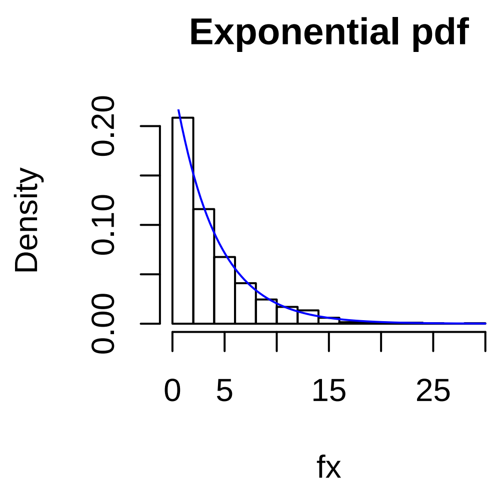
$EX = \frac{1}{\lambda} = \frac{1}{0.25} = 4$

$VX = \frac{1}{\lambda^2} = \frac{1}{(0.25)^2} = 16$

$SDX = \sqrt{VX} = 4$

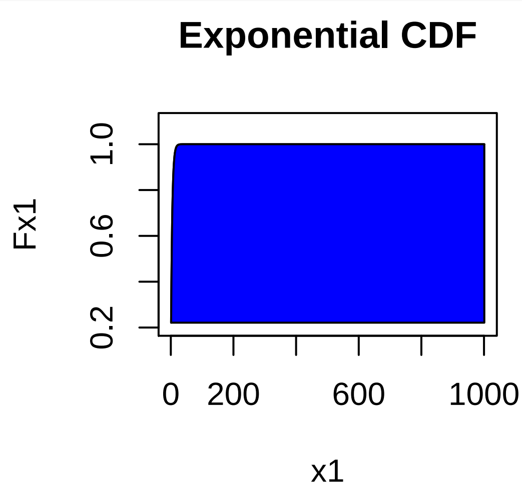
Graph of Exponential pdf

```
fx=rexp(1000,rate=.25)
hist(fx,prob=T,main="Exponential pdf")
curve(dexp(x,rate=.25),col='blue',add=T)
```



Graph of Exponential CDF

```
x1=seq(0:1000); F=antiD((1/4)*exp(-.25*x)~x); Fx1=F(x1)-F(0)
plot(x1,Fx1,type='l',main="Exponential CDF",ylim=c(0.2,1.1))
polygon(c(x1,x1[length(x1)]),c(Fx1, Fx1[1]),col='blue')
```



The Normal Distribution

The normal distribution is one of the most important and widely used. Many populations have distributions that can be fit very closely by an appropriate normal curve.

A continuous rv X is said to have a **normal distribution** with parameters μ and σ^2 where $-\infty < \mu < \infty$ and $\sigma^2 > 0$, if the pdf of X is:

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

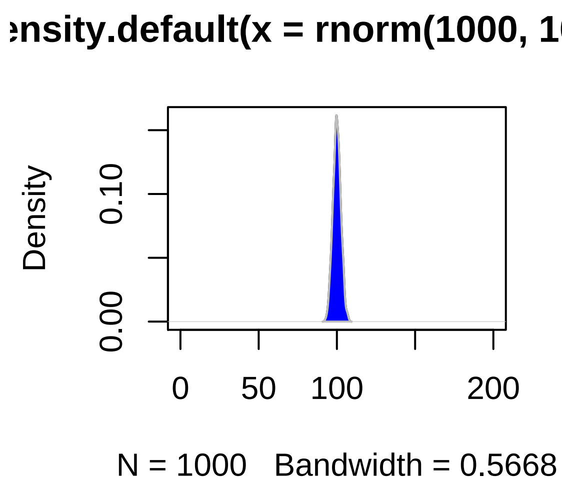
With $EX = \mu$, $VX = \sigma^2$ and $SDX = \sigma$

The Normal Density Curve

The following graphs will illustrate differences in the exact shape of the normal distribution, depending on the standard deviation (or variance). μ will be located in the center of the distribution (because of its symmetry) and σ will horizontally extend from μ to the first inflection point on the curve. Large values of σ yield graphs that are quite spread out about μ (and a value of X far from μ may be well observed), whereas small values of σ yield graphs with a high peak above μ and most of the area under the graph quite close to μ (implying that a value of X far from μ is quite unlikely).

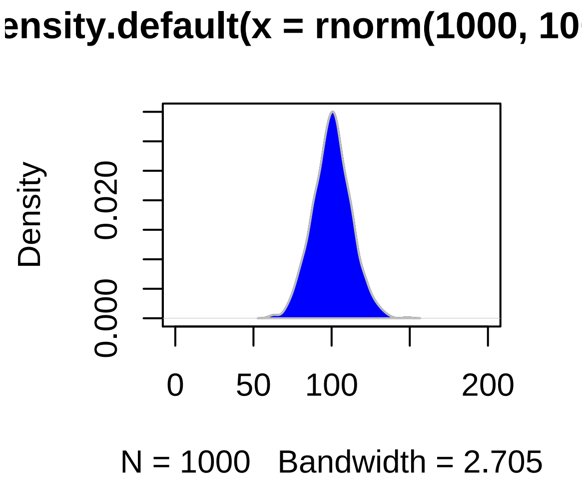
Normal Distribution mean=100, sd=2.5

```r
d1=density(rnorm(1000,100,2.5))
plot(d1,xlim=c(0,200))
polygon(d1, col="blue",border="grey")
```



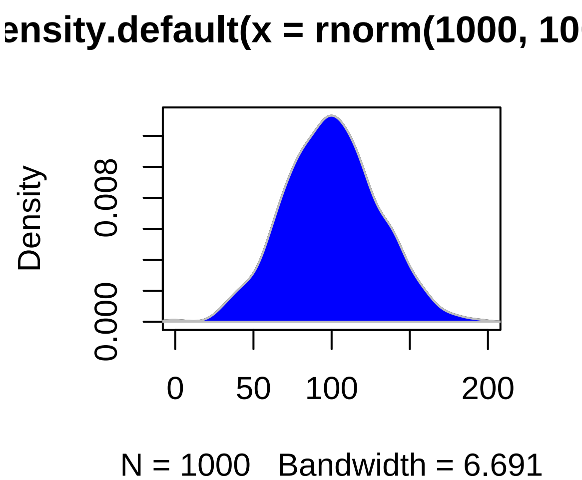
Normal Distribution mean=100, sd=12

```r
d2=density(rnorm(1000,100,12))
plot(d2,xlim=c(0,200))
polygon(d2, col="blue",border="grey")
```



Normal Distribution mean=100, sd=29

```
d3=density(rnorm(1000,100,29))
plot(d3,xlim=c(0,200))
polygon(d3, col="blue",border="grey")
```



Normal pdf

When X is a normal rv with mean μ and variance σ^2,

$$X \sim N(\mu, \sigma)$$

To compute the probabilities, this requires techniques beyond the usual methods; for $\mu = 0$ and $\sigma = 1$, tables are used, tabulated for certain values of a and b. The table is also used for any values of μ and σ by standardizing the value and using the table (or software).

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

Notation

When X is a normal rv with mean μ and variance σ^2,

$$X \sim N(\mu, \sigma)$$

$$z = \frac{X - \mu}{\sigma}$$
$$P(Z < z) = \Phi(z)$$

Standard Normal Example

With z-scores:

(1) $P(Z < 1)$
(2) $P(Z > 1)$
(3) $P(Z < -1)$
(4) $P(Z > -1)$
(5) $P(-1 < Z < 1)$
(6) z-score for top 1%
(7) z-score for $Q1$
(8) z-score for $Q3$

Normal Example I

Suppose that the diameter at breast height (in.) of trees of a certain type is normally distributed with mean 8.8 and standard deviation 2.8 ($\mu = 8.8,\ \sigma = 2.8$) ($X \sim N(8.8, 2.8)$). Calculate the following:

(1) Probability the diameter of a randomly selected tree will be at least 10“, exceed 10”
(2) Probability the diameter of a randomly selected tree will exceed 20"
(3) Probability the diameter of a randomly selected tree will be between 5" and 10"
(4) Widest 8% are wider than what diameter
(5) If four trees are selected at random, what is the probability that at least one has a diameter exceeding 10"

Normal Example II (bahaha)

Spongeboob

Standard Normal Solutions I

With z-scores and StatDistributions.com:

(1) $P(Z < 1) = \Phi(1) = 0.841345$: input z-score, left tail
(2) $P(Z > 1) = 1 - \Phi(1) = 0.158655$: input z-score, right tail
(3) $P(Z < -1) = \Phi(-1) = 0.158655$: input z-score, left tail
(4) $P(Z > -1) = 1 - \Phi(-1) = 0.841345$: input z-score, right tail
(5) $P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 0.682689$: input z-score, two tails

Standard Normal Solutions II

(6) z for top 1% is the same as the bottom 99% (or 99^{th} percentile). $z_{0.99} = 2.326348$: input 0.01 in the p-value, right tail OR 0.99 in p-value and left tail

(7) z for $Q1$ is $z_{0.25} = -0.67449$: input 0.25 in the p-value, left tail

(8) z for $Q3$ is $z_{0.75} = 0.67449$: input 0.75 in the p-value, left tail

Empirical Rule derivation:

(9) $P(-2 < Z < 2) = 0.9545$: input z-score, two tail

(10) $P(-3 < Z < 3) = 0.9973$: input z-score, two tails

Normal Solutions I part 1

(1) $P(X < 10) = P\left(Z < \frac{10-8.8}{2.8}\right) = \phi(0.43) = 0.665882$, $P(X > 10) = P\left(Z > \frac{10-8.8}{2.8}\right) = 1 - \phi(0.43) = 0.334118$

(2) $P(X > 20) = P\left(Z > \frac{20-8.8}{2.8}\right) = 1 - \phi(4) = 0.000032$

(3) $P(5 < X < 10) = P\left(\frac{10-8.8}{2.8} < Z < \frac{5-8.8}{2.8}\right) = \phi(0.43) - \phi(-1.36) = 0.578515$

Normal Soultions I part 2

(4) Find z for top 8% (same as bottom 92%). $z_{0.92} = 1.405072$. Use z-score equation and solve for x. $z = \frac{X-\mu}{\sigma} \Rightarrow x = z\sigma + \mu$. $x = (1.41)(2.8) + 8.8 = 12.748$

(5) Find the probability for one item first: $P(X > 10) = 0.334118$. Now, since they are independent, if *four* trees are selected at random, what is the probability that at least one has a diameter exceeding 10“. $P(X_4 > 1) = 1 - (X_4 = 0) = 1 - (0.334118)^4 = 0.987538$

Normal Solutions II

There is no solution for a normal Spongebob… he will be crazy forever… forever… forever…