CrossValidated

How to find pdf of a joint distribution in R?

Asked 5 years, 4 months ago Active 5 years, 4 months ago Viewed 7k times

1

$$F(x,y) = \frac{1}{6}(x^2\,y + x\,y^2)\,, \quad 0 \le x \le 2,\ 0 \le y \le 1$$

Above is the joint distribution given,

1. how to find out cumulative distribution function of y?
2. how to obtain joint probability density function of x and y?

I am a beginner in R, I know basic commands.

thanks for the help

r pdf cdf

edited Aug 18 '14 at 10:27 asked Aug 17 '14 at 17:24

user3309191

1 R doesn't magically do statistics for you. Nor is it particularly strong in symbolic algebra manipulation. If you don't know how to find a distribution, that is a statistical question, not a programming question. – MrFlick Aug 17 '14 at 17:26

1 What have you tried so far? Could you show us some code of what you've already done? – Will Beason Aug 17 '14 at 20:31

It appears your function is intended to represent a *cdf*. As such, I've replaced f with F, but if that's wrong you should alter your question (in several places) and make the exact question clear (no matter whether you read it as a mis-scaled pdf or a correctly-scaled cdf, there are things wrong with the original question). Please carefully check the question. – Glen_b -Reinstate Monica Aug 18 '14 at 10:11

This appears to be homework or otherwise to fall under the `self-study` tag. Please read the tag wiki info, add the tag if it applies and make additional edits to comply with the discussion there. – Glen_b -Reinstate Monica Aug 18 '14 at 10:12

2 Answers

This is a double integral in R (It's not done symbolically as Mathematica would do it but rather numerically):

3

```
llimy <- 0;  llimx=0
ulimy <-  2    ; ulimx=1

f <- function(x,y) 1/6*(x^2*y+x*y^2)


integrate(function(y) {
   sapply(y, function(y) {
     integrate(function(x) f(x,y), llimx, ulimx)$value
   })
 }, llimy, ulimy)
#  0.3333333 with absolute error < 3.7e-15
```

The joint pdf is just the function, f, divided by the value of the integral over the full range of the values.

answered Aug 18 '14 at 5:55

DWin

6,280 14 30

1

If you're trying to do this symbolically, you may want to try Wolfram Alpha. If you don't understand how to do this symbolically by hand, neither Wolfram Alpha nor this post will help you. You'll need to consult your statistics textbook for that. What this post does answer is how to get R to numerically compute distributions.

The volume under the curve is 1/3, so we just multiply by 3 to get the probability distribution for x and y. Obviously R doesn't deal with symbolic algebra (without the `Ryacas` package), but it is fairly easy to make pdfs and cdfs of functions. There is probably a simpler or more computationally efficient way, but this solution is fast enough for what you may be trying to do.

First, we input the pdf of `x` and `y`.

```
pdfxy <- function(x, y) (x^2 * y + x * y^2)/2
```

We convert this to a pdf of just `y` by integrating over the possible `x` values. The `sapply` function makes it so this function can easily take vectors as the `y` argument.

```
pdfy <- function(y) {
  result <- sapply(y, function(b) integrate(function(a) pdfxy(a, b), 0, 2)$value)
  return(result)
}
```

We then make this into a cdf by integrating over `y` from 0 to the desired value. The sapply function in this function is only required if you want to be able to input vectors.

```
cdfy <- function(y) {
  result <- sapply(y, function(a) integrate(function(b) pdfy(b), 0, a)$value)
  return(result)
}
```

Now we can just type

```
> cdfy(c(0,0.5,1))
[1] 0.0000000 0.2083333 1.0000000
> cdfy(0.4)
[1] 0.128
```

You now have a function in R that calculates the cumulative distribution of y given any y.

answered Aug 17 '14 at 21:25
Will Beason

That's not the answer you get if you go to Wolfram Alpha and get the indefinite integral: x^3*y/3 + x^2*y^2/2 over those limits is 4.66667. (Plus is doesn't make sense to have only one limit of integration for this problem.) – DWin Aug 18 '14 at 1:47

@BondedDust You should check your math: bit.ly/1vZiX8v. Remember that you need to integrate over both variables, not just x. – Will Beason Aug 18 '14 at 2:01

That wasn't the difference. My definite integral was not divided by 6. Now, do it with R. That's where I think you are not addressing the question. – DWin Aug 18 '14 at 5:03