

**Class Example 1. A Discrete Distribution**

Define the discrete distribution with probability mass function,

$$p(k) = e^{-3} \frac{3^k}{k!}, \quad \text{for } k = 0, 1, 2, \dots$$

(a) It can be shown that,

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}.$$

Use this to argue now that  $p(\cdot)$  is a probability mass function.

**Solution:** It is clear that  $p(k) \geq 0$  and further the probabilities add up to 1:

$$\sum_{k=0}^{\infty} p(k) = e^{-3} \sum_{k=0}^{\infty} \frac{3^k}{k!} = e^{-3} e^3 = 1.$$

(b) Plot the probability mass function, for  $k = 0, 1, 2, \dots, 10$ .

```
using PyPlot
kArray = collect(0:10)
pArray = [e^(-3) * 3^k/factorial(k) for k in kArray]
PyPlot.stem(kArray, pArray);
```

(c) It is well known that the mean of this (Poisson distribution) is 3. Use Julia to compute,

$$\sum_{k=0}^{10} k p(k) \approx \sum_{k=0}^{\infty} k p(k) = 3.$$

```
sum(kArray .* pArray)
```

(d) It is also well known that the variance is also 3 (for this type of distribution the mean equals the variance). Use Julia to compute,

$$\sum_{k=0}^{10} (k - 3)^2 p(k) \approx 3.$$

```
sum((kArray .- 3) .^2 .* pArray)
```

**Class Example 2. A Continuous Distribution**

Consider a continuous distribution with probability density function (pdf) on  $[-2, 2]$ ,

$$f(x) = \begin{cases} \frac{2}{3} + \frac{1}{3}x, & x \in [-2, -1], \\ \frac{1}{3}, & x \in (-1, 1), \\ \frac{2}{3} - \frac{1}{3}x, & x \in [1, 2]. \end{cases}$$

Assume that the random variable  $X$  is distributed according to  $f(x)$ .

- (a) Plot the PDF of  $X$  in Julia.

```
using PyPlot
function f(x)
    if -2 <= x <= -1
        return 2/3 + x/3
    elseif -1 < x < 1
        return 1/3
    elseif 1 <= x <= 2
        return 2/3 - x/3
    else
        return 0
    end
end
x = linspace(-2,2,1000)
y = [f(u) for u in x]
PyPlot.plot(x, y)
xlabel("x")
ylabel("f(x)")
title("The PDF, f(x)");
```

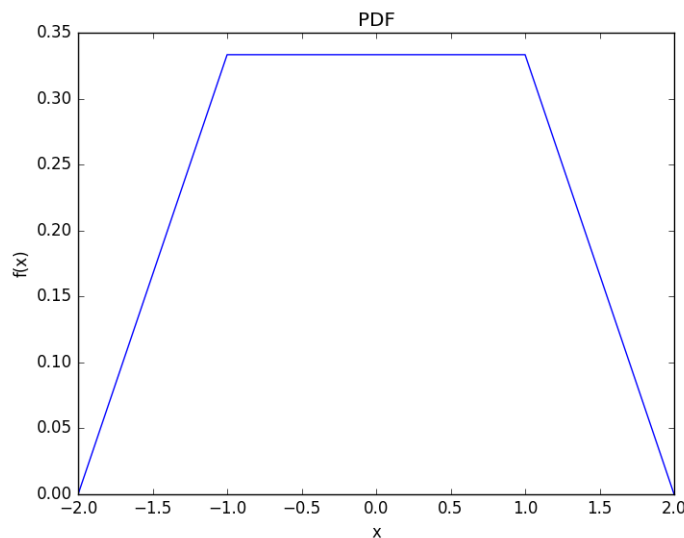


Figure 1: The pdf,  $f(x)$ .

- (b) Plot the CDF of  $X$  in Julia.

In order to do this, we will numerically integrate the PDF. Remember that the density  $f(x)$  has the following meaning for small  $\Delta x$ :

$$f(x) \Delta x \approx P(X \in [x, x + \Delta x]).$$

Hence,

$$F(x) = P(X \leq x) = \int_{-2}^x f(u) du \approx \sum_{u=-2}^x f(u)\Delta x.$$

We can carry out the sum on the right in Julia. Here we choose  $\Delta x = 0.001$ :

```
function F(x)
    sum([f(u)*0.001 for u in -2:0.001:x])
end
x = linspace(-2.5,2.5,1000)
y = [F(u) for u in x]
PyPlot.plot(x, y)
xlabel("x")
ylabel("F(x)")
title("The CDF, F(x)")
```

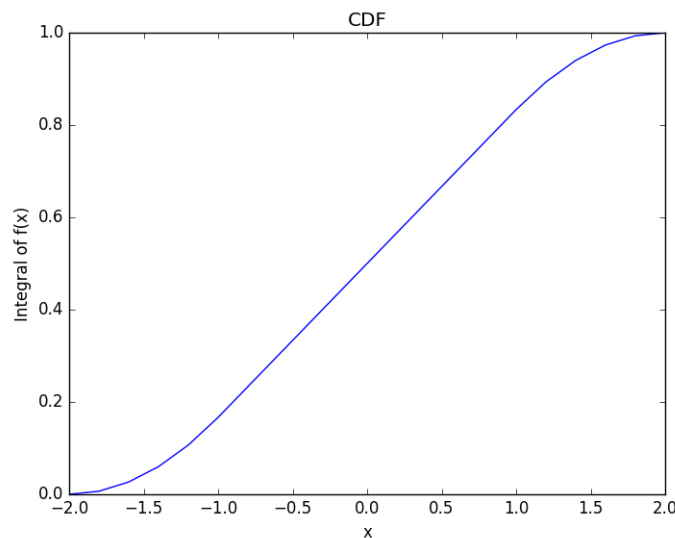


Figure 2: The CDF  $F(x)$ .

Note that here we numerically integrated  $f(x)$ , but we can also do it analytically in this case. This would require integrating each part of  $f(x)$  separately.

(c) Looking at the distribution, what is its mean?

- (i) Argue why the mean is 0 by looking at the PDF.
- (ii) Calculate the mean analytically.

**Solution:** We see that  $f(x) = f(-x)$  (it is a symmetric function). Now,

$$\int_{-2}^2 x f(x) dx = \int_{-2}^0 x f(x) dx + \int_0^2 x f(x) dx.$$

Now changing variables in the first integral ( $u = -x$ ) we get,

$$= \int_0^2 (-u) f(-u) du + \int_0^2 x f(x) dx.$$

Now using the symmetry of  $f$ ,

$$= - \int_0^2 u f(u) du + \int_0^2 x f(x) dx = 0.$$

**Class Example 3. Common Families of Distributions**

We now explore the DiscreteUniform, Binomial, Exponential and Normal distributions using Julia. For each such distribution, do the following:

- (a) Create a variable representing the distribution with some parameters of your choice (within the valid parameter range). Here is an example of a Normal distribution with  $\mu = 2$  and  $\sigma = 0.5$

```
using Distributions
normDist = Normal(2,0.5)
```

- (b) Use `mean()` to obtain the mean value. Compare it to the value as calculated using the formula in the course material.

```
mean(normDist) , 2
```

- (c) Generate 100,000 random variables from the distribution and calculate the sample mean. Compare to (b) to ensure the values are close.

```
randomSample = rand(normDist,10^5)
mean(randomSample), sum(randomSample)/10^5
```

Note that in the cell above the function `mean` performs the equivalent of `sum(...)/n`. However when `mean` is performed on a **Distribution** type object (**a**), it does not perform this sum calculation, but rather returns the mean defined by the Distribution in question. In Julia, this type of feature is called multiple dispatch.

- (d) Plot a histogram of the distribution taking care to present discrete and continuous distributions in an appropriate manner.

```
PyPlot.plt[:hist](rand(normDist,10^5),100);
```

**Solution:**

We now carry this out for all four families of distributions:

We first create an array of distributions. We then create a matrix of means for these distributions using three different methods.

We then plot the PMF/PDF of these four distributions over pre-defined domains, and finally we plot histograms of these four distributions, using a Monte Carlo sampling method.

```
using PyPlot, Distributions

dists = [DiscreteUniform(-3,4),
         Binomial(20,0.25),
         Exponential(3),
         Normal(1.5,0.25)];

meansFromFormula = [ (-3+4)/2, 20*0.25, 3, 1.5];
meansFromDistributionsPackage = [mean(d) for d in dists]

function estimateMean(dist)
    mean(rand(dist,10000))
end

meansFromMonteCarlo = [estimateMean(d) for d in dists]

meansMatrix = [ ["DiscreteUniform(-3,4)", "Binomial(20,0.25)",
                 "Exponential(3)", "Normal(1.5,0.25)"]'; meansFromFormula';
                meansFromDistributionsPackage'; meansFromMonteCarlo']'

support = [-3:4,
           0:20,
           linspace(0,20,1000),
           linspace(0.5,2.5,1000)]

subplot(421)
PyPlot.stem(support[1],pdf(dists[1],support[1]))

subplot(422)
PyPlot.stem(support[2],pdf(dists[2],support[2]))

subplot(423)
PyPlot.plot(support[3],pdf(dists[3],support[3]))

subplot(424)
PyPlot.plot(support[4],pdf(dists[4],support[4]))

base = 424
for i in 1:4
    subplot(base+i)
    PyPlot.plt[:hist](rand(dists[i],10^5),100);
end

meansMatrix
```

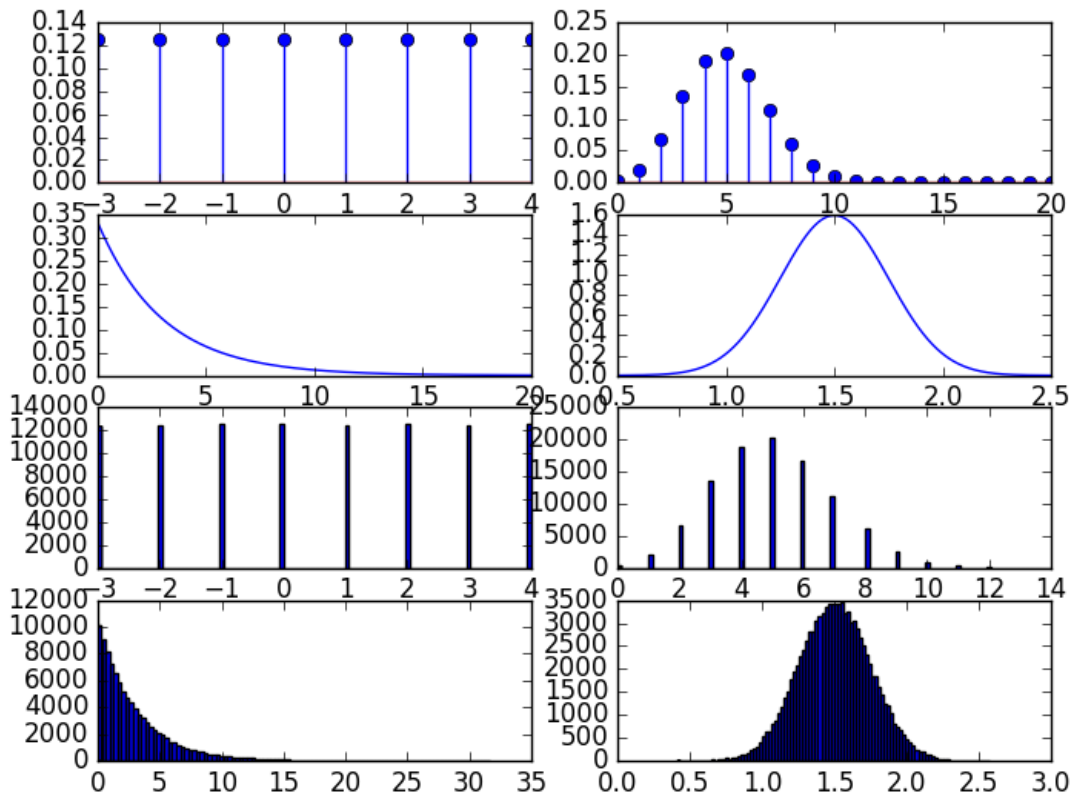


Figure 3: Plot of distributions and histograms

**Question 1. A Discrete Distribution – PMF**

Verify that  $p(x)$  is a probability mass function (pmf) and calculate the following for a random variable  $X$  with this pmf:

$x$	1.25	1.5	1.75	2	2.25
$p(x)$	0.2	0.4	0.1	0.2	0.1

- $P(X \geq 2)$
- $P(X > 1.65)$
- $P(X = 1.5)$
- $P(X < 1.3 \text{ or } X > 21)$
- The mean.
- The variance.
- Sketch the cumulative distribution function (cdf). Note that it exhibits jumps and is a right continuous function.

**Question 2. A Discrete Distribution – CDF**

Given the cdf,  $F(x)$  below for the random variable,  $X$ , calculate the following:

$$F(x) = \begin{cases} 0 & x < -10 \\ 0.25 & -10 \leq x < 30 \\ 0.75 & 30 \leq x < 50 \\ 1 & 50 \leq x \end{cases}$$

- $P(X \leq 50)$
- $P(X \leq 40)$
- $P(40 \leq X \leq 60)$
- $P(X < 0)$
- $P(0 \leq X < 10)$
- $P(-10 < X < 10)$
- The mean.
- The standard deviation.
- Sketch the probability mass function (PMF).

**Question 3. Guessing on Multiple Choice Exams**

A multiple-choice test contains 25 questions, each with four answers. Assume that a student just guesses on each question.

- What is the probability that the student answers more than 20 questions correctly?
- What is the probability that the student answers fewer than 5 questions correctly?

The following code generates the vector `pmfValues` from a Binomial distribution with parameters  $n = 10$  and  $p = 0.4$ . It then sums up the vector, illustrating that the sum of all of the probabilities is 1.

```
using Distributions
bDist = Binomial(10,0.4)
pmfValues = [pdf(bDist, x) for x in 0:20]
sum(pmfValues[1:11])
```

(c) Modify the code above, to validate your answers in (a) and (b).

#### Question 4. Stuck in Traffic

A particularly long traffic light on your morning commute is green 20% of the time that you approach it. Assume that each morning represents an independent trial.

- Over 5 mornings, what is the probability that the light is green on exactly one day?
- Over 20 mornings, what is the probability that the light is green on exactly four days?
- Over 20 mornings, what is the probability that the light is green on more than four days?
- What is the mean number of days with green light during a month of 30 days?

Optional: You may verify your analytic answers using Julia, in a similar manner to Question 3c.

#### Question 5. Aerospace Inspections

The thickness of a flange on an aircraft component is Uniformly distributed between 0.95 and 1.05 millimetres. Determine the following:

- Cumulative distribution function of flange thickness.
- Proportion of flanges that exceeds 1.02 millimetres.
- Thickness exceeds 90% of the flanges.
- Mean and variance of flange thickness.
- Assume now that you are sampling 10 independent flanges. What is the variance of the number of flanges with a thickness less than 0.97 millimetres?

#### Question 6. Mobile Phone Semiconductors

The line width for semiconductor manufacturing is assumed to be Normally distributed with a mean of 0.5 micrometers and a standard deviation of 0.05 micrometers.

- What is the probability that a line width is greater than 0.62 micrometer?
- What is the probability that a line width is between 0.47 and 0.63 micrometer?
- The line width of 90% of samples is below what value?

#### Question 7. The Prototype Shoe

The weight of a sophisticated running shoe is normally distributed with a mean of 12 ounces and a standard deviation of 0.5 ounce.

- What is the probability that a shoe weighs more than 13 ounces?



- (b) What must the standard deviation of weights be in order for the company to state that 99% of its shoes weighs less than 13 ounces?
- (c) If the standard deviation remains at 0.5 ounce, what must the mean weight be for the company to state that 99% of its shoes weighs less than 13 ounces?

**Question 8. Time Until (Blue Screen of Death) BSoD**

Suppose that the time to failure (in hours) of hard drives in a personal computer can be modelled by an exponential distribution with  $\lambda = 0.0003$ .

- (a) What proportion of the hard drives will last at least 10,000 hours?
- (b) What proportion of the hard drives will last at most 7,000 hours?
- (c) What is the variance of the time until failure for a hard drive?
- (d) Use Monte Carlo simulation to predict the following: Assume a computer now has two independent hard-drives and the failure of the computer occurs once both hard-drives have died. What is the mean life of the computer?