MATH7501 Practical 11 (Week 12), Semester 1-2021

Topic: Probability Distributions

Author: Aminath Shausan
Date: 21-05-2021

Pre-Tutorial Activity

- Students must have familiarised themselves with units 9 and 10 contents of the reading materials for MATH7501

Resources

- Chapters 9 and 10: Computation of mean, variance, expectation, gradient decent method
- https://en.wikipedia.org/wiki/Rayleigh_distribution

Section 1: The Rayleigh Distribution

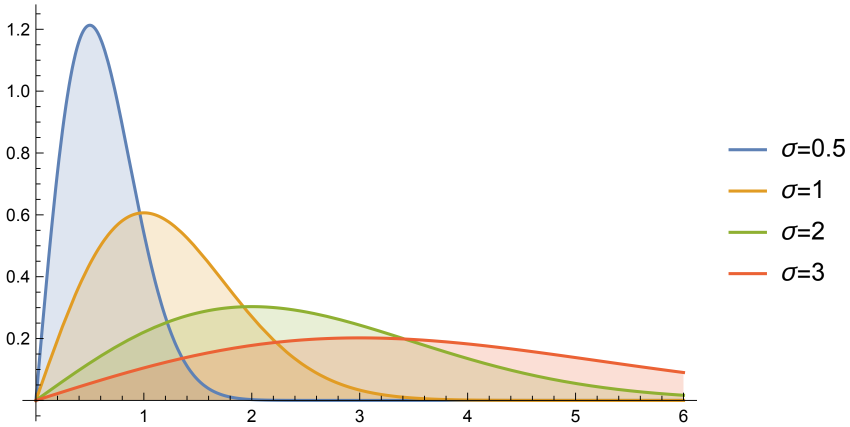
In probability theory and statistics,
the **Rayleigh distribution** is a continuous probability distribution for nonnegative – valued random variables.

The notation X ~ Rayleigh (σ) means that the random variable X has a Rayleigh distribution with shape parameter σ. The probability density function (pdf) is :

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}, & x \ge 0, \ \sigma > 0 \\ 0, & x < 0 \end{cases}$$

Q1) Plot the pdf of X for σ = 0.5, 1, 2, 3

In[•]:=
```
Plot[Table[PDF[RayleighDistribution[σ], x], {σ, {.5, 1, 2, 3}}] // Evaluate,
 {x, 0, 6}, Filling -> Axis, PlotRange → All,
 PlotLegends → {"σ=0.5", "σ=1", "σ=2", "σ=3"}]
```

Out[•]=


Q2) Show that the cumulative distribution function (cdf) of X is

$$F_X(x) = \begin{cases} 1 - e^{\frac{-x^2}{2\sigma^2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$F_X(x) = P(X \leq x) = \int_0^x f_X(t)\,dt$$
$$= \int_0^x \frac{t}{\sigma^2} e^{\frac{-t^2}{2\sigma^2}}\,dt$$

Let $u = \frac{t^2}{2\sigma^2}$. Then $du = \frac{t}{\sigma^2}\,dt$,

and the limits of the integral are $u = 0$ and $u = \frac{x^2}{2\sigma^2}$. Thus,

substituting $dt = \frac{\sigma^2}{t}du$ in the above integration, we have,

$$F_X(x) = \int_0^{\frac{x^2}{2\sigma^2}} e^{-u}\,du$$
$$= -\left[e^{-u}\right]_{u=0}^{u=\frac{x^2}{2\sigma^2}}$$
$$= -\left\{e^{-\frac{x^2}{2\sigma^2}} - e^0\right\}$$
$$= \begin{cases} 1 - e^{-\frac{x^2}{2\sigma^2}}, & x \geq 0 \\ 0, & x < 0, \end{cases}$$

as required.

You can also use the Integrate[] function to compute the cdf as shown below.

In[•]:= $f[x\_] := \frac{x}{\sigma^2} E^{-\frac{x^2}{2\sigma^2}}$ (*Rayleigh probability density function*)

```
Integrate[f[x], {x, 0, u}] (*cdf*)
```

Out[•]= $1 - e^{-\frac{u^2}{2\sigma^2}}$

Q3) Show that $f_X(x)$ is a valid probability density function by showing that the integral over $[0, \infty)$ is 1

Method 1 : Show this using calculus

$$\int_0^\infty f_X(x)\,dx = \int_0^\infty \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}\,dx$$

$$= \text{limit}_{y\to\infty} \int_0^y \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}\,dx$$

$$= \text{limit}_{y\to\infty} F_X(y),\ \text{by the definition of the cdf}$$

$$= \text{limit}_{y\to\infty} \left(1 - e^{-\frac{x^2}{2\sigma^2}}\right)$$

$$= 1 - 0$$

$$= 1,\ \text{as required.}$$

Method 2 :

Use the Integrate[] function to compute this integration exactly in Mathematica.

In[∘]:=
```
Integrate[f[x], {x, 0, ∞}, Assumptions → σ > 0]
```

Out[∘]= 1

Method 3 : You can also use NIntegrate[]

function to derive a numerical approximation to $\int_0^\infty \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}\,dx$,

for a given value of σ (say for example, $\sigma = 1$).

```
NIntegrate[x E^(-x^2/2), {x, 0, ∞}]
```

Out[∘]= 1.

Method 4 :

Approximate the integral by Reiman sum. That is $\int_0^\infty f_X(x)\,dx \simeq \sum_{k=0}^{\infty} f(k)\,\delta$,

where δ is the width of the rectangles.

(*discretisation sum *)

In[14]:=
```
Clear[δ] (*δ is the width of the rectangle*)
```

In[20]:=
```
Total[Table[ x E^(-x^2/2) δ, {x, 0, 1000, δ = 0.001}]]
```

Out[20]= 1.

Q4) Show that the mean of X ~ Rayleigh (σ) $= \sigma\sqrt{\frac{\pi}{2}}$

$$\mu_X = E(X) = \int_0^\infty x f_X(x)\,dx$$

$$= \int_0^\infty x \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}\,dx$$

Let $u = \frac{x^2}{2\sigma^2}$. Then $du = \frac{x}{\sigma^2}dx$,

and the limits of the integral are still $u = 0$ and $u = \infty$. Thus,

substituting $dx = \frac{\sigma^2}{x}du$ and using $x = \sqrt{(2\sigma^2 u)}$ in the above integration we have,

$$\mu_X = \int_0^\infty \sqrt{(2\sigma^2 u)}\, e^{-u}\, du$$
$$= \sigma\sqrt{2}\int_0^\infty \sqrt{u}\, e^{-u}\, du$$
$$= \sigma\sqrt{2}\int_0^\infty u^{3/2-1}\, e^{-u}\, du$$
$$= \sigma\sqrt{2}\,\Gamma\left(\frac{3}{2}\right),\ \Gamma(x) \text{ is the Gamma function defined as : } \Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$
$$= \sigma\sqrt{2}\,\Gamma\left(\frac{1}{2}+1\right)$$
$$= \sigma\sqrt{2}\times\frac{1}{2}\Gamma\left(\frac{1}{2}\right),\ \text{using the property of Gamma function, } \Gamma(x+1) = x\Gamma(x)$$
$$= \sigma\sqrt{2}\times\frac{1}{2}\sqrt{\pi},\ \text{using the property of Gamma function, } \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$
$$= \sigma\sqrt{\left(\frac{\pi}{2}\right)}$$

You can use the Integrate[] function to check your derivation of the mean as follows.

In[∘]:= `Integrate[x f[x], {x, 0, ∞}, Assumptions → σ > 0] (*mean (μ) *)`

Out[∘]= $\sqrt{\frac{\pi}{2}}\ \sigma$

In[∘]:= `(*using numerical approximation to compute the mean when σ =1*)`

In[∘]:= NIntegrate$\left[x\ x\ E^{-\frac{x^2}{2}}, \{x, 0, \infty\}\right]$

Out[∘]= 1.25331

`δ = 0.01`

Out[∘]= 0.01

In[∘]:= Total$\left[\text{Table}\left[x\ x\ E^{-\frac{x^2}{2}}\ \delta, \{x, 0, 10, \delta\}\right]\right]$

Out[∘]= 1.25331

Q5) Show that the variance of $X \sim$ Rayleigh $(\sigma) = \sigma^2 \left(\frac{4-\pi}{2}\right)$

From Week 11 prctical, we know that $\sigma^2{}_X = \text{var}(X) = E[X^2] - \mu_X{}^2$,
where $E[X^2]$ is the second moment of X and μ_X is the mean of X,
which we computed in Q4. Thus,
it remains to find an expression for $E[X^2]$. We will require the integration bi - parts formula $\int u\, dv = uv - \int v\, du$ for this calculation.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$$

$$= \int_0^{\infty} x^2 \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} dx$$

$$= \left[x^2 \int_0^{\infty} \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} dx\right]_{x=0}^{x=\infty} - \int_0^{\infty} 2x \left(\int_0^{\infty} \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} dx\right) dx,$$

by using $u = x^2$ and $dv = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}$,

$$= \left[-x^2 e^{\frac{-x^2}{2\sigma^2}}\right]_{x=0}^{x=\infty} + \int_0^{\infty} 2x\, e^{\frac{-x^2}{2\sigma^2}} dx$$

To evaluate the integral, let $u = \frac{x^2}{2\sigma^2}$. Then $du = \frac{x}{\sigma^2} dx$,
and the limits of the integral are still $u = 0$ and $u = \infty$. Thus,
substituting $dx = \frac{\sigma^2}{x} du$ we have

$$E[X^2] = 0 + 2\sigma^2 \left[\int_0^{\infty} e^{-u} du\right]$$
$$= -2\sigma^2 \left[e^{-u}\right]_{u=0}^{u=\infty}$$
$$= -2\sigma^2 [0 - 1]$$
$$= 2\sigma^2$$

Thus, $\sigma^2{}_X = \text{var}(X) = E[X^2] - \mu_X{}^2$

$$= 2\sigma^2 - \left(\sigma \sqrt{\left(\frac{\pi}{2}\right)}\right)^2$$

$$= \sigma^2 \left(\frac{4-\pi}{2}\right) \text{ as required.}$$

You can use the Integrate[] function to check your work.

```
Integrate[x^2 f[x], {x, 0, ∞}, Assumptions → σ > 0]
(*compute second moment (E[X^2]*)
```

Out[•]= $2\sigma^2$

```
(*using numerical approximation to compute the second moment when σ =1*)
```

```
In[•]:= NIntegrate[x^2 x E^(-x^2/2), {x, 0, ∞}]
```

Out[•]= 2.

```
In[•]:= Total[Table[x^2 x E^(-x^2/2) δ, {x, 0, 10, δ}]]
```

Out[•]= 2.

In[•]:= (*variance*)

$$E\left[X^2\right] - \mu^2 = 2\sigma^2 - \left(\sqrt{\frac{\pi}{2}}\ \sigma\right)^2$$

Out[•]= $2\sigma^2 - \frac{\pi\sigma^2}{2}$

Q6) Find the median of X.

Note that the median of X is the number M such that,

$$\int_0^M f_X(x)\,dx = \frac{1}{2}$$

Note that the left hand side is the cdf of X evaluated at M. Thus we have

$$F_X(M) = \frac{1}{2}$$
$$1 - e^{-\frac{M^2}{2\sigma^2}} = \frac{1}{2}$$
$$e^{-\frac{M^2}{2\sigma^2}} = \frac{1}{2}$$
$$-\frac{M^2}{2\sigma^2} = \ln\left(\frac{1}{2}\right)$$
$$M^2 = -2\sigma^2 \ln\left(\frac{1}{2}\right)$$

$M = \sigma\sqrt{\left(-2\ln\left(\frac{1}{2}\right)\right)}$ as the median.

Q7) The quantile function of the distribution, q (u) for $u \in [0, 1)$,
is defined as follows : For each u, we should have,

$$\int_0^{q(u)} f_X(x)\,dx = u$$

a) Find an expression for q (u)

By noting that the left -
hand side of the equation is the cdf of X evaluated at q (u), we have

$$F_X(q(u)) = u$$
$$1 - e^{-\frac{q(u)^2}{2\sigma^2}} = u$$
$$e^{-\frac{q(u)^2}{2\sigma^2}} = 1 - u$$
$$-\frac{q(u)^2}{2\sigma^2} = \ln(1-u)$$
$$q(u)^2 = -2\sigma^2\ln(1-u)$$
$$q(u) = \sigma\sqrt{(-2\ln(1-u))}$$

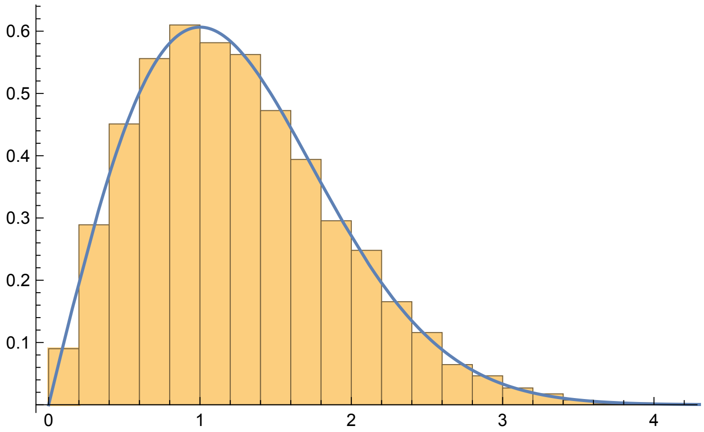
b) Say that $X = q(U)$ with U as Uniformly distributed on [0, 1]. Then,
X has a Rayleigh distribution. Show this empirically for $\sigma = 1$,
by generating 10^4 uniform random variables on [0, 1]

```
(*Generate a Rayleigh random Variable from Uniform[0,1] random variable*)
```

```
In[•]:= U = RandomReal[1, 10 000];
        X = Sqrt[-2 Log[1 - U]];

In[•]:= Show[Histogram[X, Automatic, "PDF"],
         Plot[PDF[RayleighDistribution[1], x],
          {x, 0, 8}, PlotRange → All]]
```

Out[•]=



The plot shows histogram of the data X along with the pdf of X ~ RayleighDistribution (1).

Section 2: Simple Linear Regression Problem

Consider the simple linear regression problem with data points (x_1, y_1), ..., (x_n, y_n). The aim is to seek β_0 and β_1 to fit the line,

$$y = \beta_0 + \beta_1 x,$$

by minimizing the loss function

$$L(\beta_0\ \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 .$$

Here, β_0 and β_1 are the slope and the intercept, respectively, of the line of best fit.

Q1) Compute an expression for the gradient of $L(\beta_0\ \beta_1)$

Since the loss function has two variables, we need to use partial derivatives here.

$$\begin{aligned}\frac{\partial L\ (\beta_0\ \beta_1)}{\partial \beta_1} &= \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1\ x_i))^2 \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} (y_i - (\beta_0 + \beta_1\ x_i))^2 \\ &= -2 \sum_{i=1}^{n} x_i\ (y_i - (\beta_0 + \beta_1\ x_i))\end{aligned}$$

$$\begin{aligned}\frac{\partial L\ (\beta_0\ \beta_1)}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1\ x_i))^2 \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_0} (y_i - (\beta_0 + \beta_1\ x_i))^2 \\ &= -2 \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1\ x_i))\end{aligned}$$