

MATHEMATICAL STATISTICS WITH MAPLE

ALEX POTAPCHIK

ABSTRACT. Widely adopted by universities, colleges, research institutions, and companies, Maple delivers the comprehensive environment and support resources to visualize and explore mathematical concepts, develop application tools, and share mathematical information and documents through the Web. Combined with hundreds of free add-on packages and applications available through the Maple application centre, Maple constitutes a flexible analytical tool for professors, researchers, and students. The goal of my talk is to demonstrate some of the benefits of using Maple for teaching mathematical statistics. I will focus on the new Statistics package.

The Statistics package - introduced with Maple 10 - is a collection of functions and interactive tools for mathematical statistics and data analysis. It supports a wide range of common statistical tasks, such as quantitative and graphical data analysis, simulation, and curve fitting. In contrast to purely numerical/graphical systems such as SPSS, JMP, Gauss, the package was designed to take advantage of Maple's symbolic power. The tools available in the Statistics package can greatly facilitate various tasks that are of primary interest in mathematical statistics such as calculations with abstract random variables and symbolic manipulation of likelihood functions.

1. PROBABILITY CALCULATIONS

At the most basic level, the Statistics package can be used as a reference tool to get information about various continuous and discrete probability density functions.

Example 1. The following Maple command will return the mean of a random variable which has a Weibull distribution with scale parameter b and shape parameter c .

```
> Mean(Weibull(b, c));
```

This yields

$$b \cdot \Gamma\left(\frac{c+1}{c}\right).$$

Here is the standard deviation:

Maplesoft, Waterloo, Canada.

```
> StandardDeviation(Weibull(b, c));
```

$$\sqrt{b \left(\Gamma \left(\frac{c+2}{c} \right) - \left(\Gamma \left(\frac{c+1}{c} \right) \right)^2 \right)}$$

(Note that the latex code for the above output was also generated by Maple).

We can also get the corresponding probability density function

```
> f := PDF(Weibull(b, c), t);
```

$$\begin{cases} 0 & t < 0 \\ ct^{-1+c} e^{-\left(\frac{t}{b}\right)^c} (b^c)^{-1} & \textit{otherwise} \end{cases},$$

and evaluate it at various points

```
> eval(f, t = 1);
```

$$\frac{ce^{-(b^{-1})^c}}{b^c}$$

```
> eval(f, [t = 1, c = 1/6, b = 1/3]);
```

$$1/6 \sqrt[6]{3} e^{-\sqrt[6]{3}}$$

Note that Maple stores only the most important facts about any given distribution in the data base. Everything else is calculated on the fly.

In the next set of examples we will define an abstract random variable and calculate some quantities of interest.

Example 2. The following command defines a random variable $X \equiv Normal(a, b)$, where a and b are parameters.

```
> X := RandomVariable(Normal(a, b));
```

X

Note, that X is a perfectly valid Maple object. It can be used as an argument to Maple functions and can appear as a part of larger expressions. Just as a simple test, let's compute the mean and the variance of X

```
> Mean(X), Variance(X);
```

$$a, \quad b^2.$$

Here is the 5th moment of X

> Moment(X, 5);

$$15b^4a + 10b^2a^3 + a^5$$

We can also calculate the mean of X^2 :

> Mean(X^2);

$$a^2 + b^2$$

For that matter, we can also calculate the whole density function of X^2

> PDF(X^2, t);

$$\begin{cases} 0 & t \leq 0 \\ 1/4\sqrt{2} \left(e^{-1/2 \frac{(\sqrt{t}+a)^2}{b^2}} + e^{-1/2 \frac{(-\sqrt{t}+a)^2}{b^2}} \right) \frac{1}{\sqrt{t}} \frac{1}{\sqrt{\pi}} b^{-1} & 0 < t \end{cases}$$

Similarly, we compute the density of $((X - a)/b)^2$ and compare the result with the χ^2 -density.

> PDF(((X-a)/b)^2, t) = PDF(ChiSquare(1), t);

$$\begin{cases} 0 & t \leq 0 \\ 1/2 \frac{\sqrt{2}e^{-1/2t}}{\sqrt{\pi}\sqrt{t}} & 0 < t \end{cases} = \begin{cases} 0 & t < 0 \\ 1/2 \frac{\sqrt{2}e^{-1/2t}}{\sqrt{\pi}\sqrt{t}} & otherwise \end{cases}$$

Example 3. As a little more complicated example, consider five random variables X_i each having a standard normal distribution. We will assume our random variables to be independent.

> X := seq(RandomVariable(Normal(0, 1)), i = 1..5);

Here is the density function for their sum:

> PDF(X[1]+X[2]+X[3]+X[4]+X[4]+X[5], t);

$$1/10 \frac{e^{-1/10 t^2} \sqrt{5}\sqrt{2}}{\sqrt{\pi}}$$

and their sum of squares

> PDF(X[1]^2+X[2]^2+X[3]^2+X[4]^2+X[4]^2+X[5]^2, t);

$$\begin{cases} 0 & t \leq 0 \\ 1/6 \frac{t^{3/2} \sqrt{2} e^{-1/2 t}}{\sqrt{\pi}} & 0 < t \end{cases}$$

Again, it may be instructive to compare the result with a χ^2 density with 5 degrees of freedom:

```
> PDF(ChiSquare(5), t);
```

$$\begin{cases} 0 & t \leq 0 \\ 1/6 \frac{t^{3/2} \sqrt{2} e^{-1/2 t}}{\sqrt{\pi}} & 0 < t \end{cases}$$

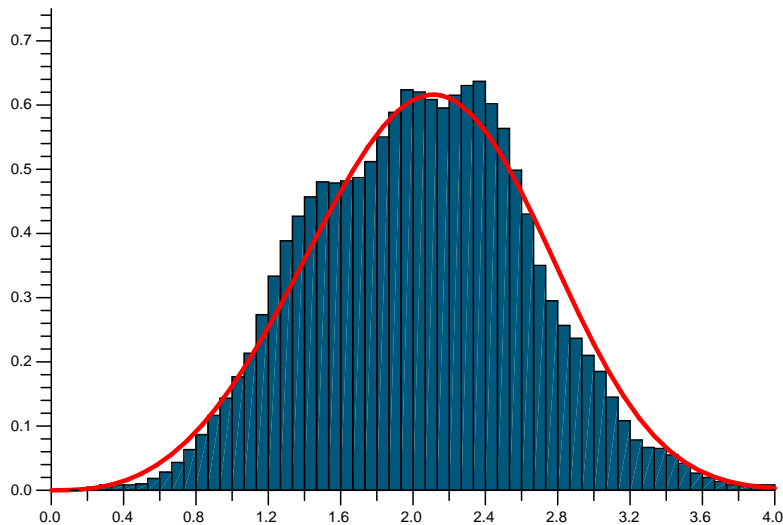
2. REGRESSION AND PARAMETER ESTIMATION

The following set of examples demonstrates the use of Maple's symbolic capabilities for parameter estimation.

Let us create a sample data set.

```
> S := Sample(Weibull(2.3, 3.7), 1000);
```

Here is how our data looks like.



We would like to fit a Weibull distribution to this data set.

One of the simplest methods is the method of moments. Since we have two parameter in our distribution, we need two equations to calculate them. So we will match the corresponding moments of the data set and the distribution.

Here is the first equation.

```
> E[1] := Moment(Weibull(a, b), 1) = Moment(S, 1);
```

$$a\Gamma\left(\frac{b+1}{b}\right) = 2.043681895$$

Here is the second equation.

```
> E[2] := Moment(Weibull(a, b), 2) = Moment(S, 2);
```

$$a^2\Gamma\left(\frac{b+2}{b}\right) = 4.532388451$$

Now all we need is to solve these equations.

```
> fsolve({E[1], E[2]}, {a, b});
```

We thus obtain our first solution.

$$\{a = 2.260287371, b = 3.828867240\}$$

We can refine the estimate using the maximum likelihood method. The procedure is very simple. We build the likelihood function symbolically and find the values of a and b that maximize the value of this function.

```
> F := LogLikelihood(Weibull(a, b), S);
> NLPsolve(-F, a = 2.1 .. 2.4, b = 3.5 .. 3.8);
```

This gives our second solution

$$[901.045550122, [a = 2.261386112, b = 3.788696941]]$$

Yet another option is to use rank regression. First we will rank our data set and convert the ranks to probabilities.

```
> A := map(t -> 0.001 t, Rank(S));
> f := CDF(Weibull(a, b), t);
> g := NonlinearFit(f, S, A, t);
```

To find a and b we simply solve the equations.

```
> solve({eval(f = g, t = 2), eval(f = g, t = 1)}, {a, b});
```

This gives our third solution.

$$\{a = 2.255897938, b = 3.776849178\}$$

3. RANDOM VARIATE GENERATION / SIMULATION

The package provides optimized algorithms for simulating from all supported distributions as well as tools for creating custom random number generators.

Example 1. The following command will generate a random sample drawn from the standard normal distribution.

```
> Sample(Normal(0, 1), 10^6);
```

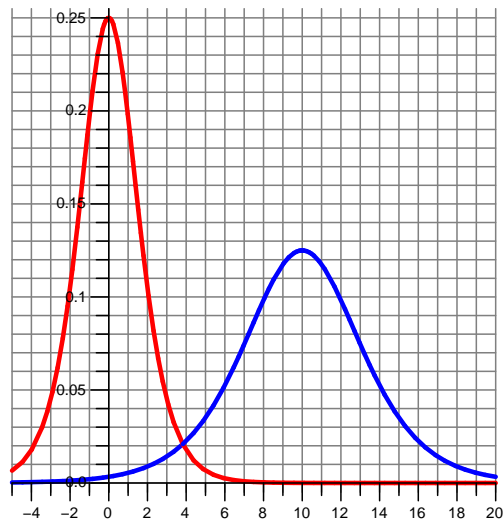
Alternatively, the Sample command can return a Maple procedure which can be used to generate random samples drawn from the normal distribution.

```
> F := Sample(Normal(0, 1));
> F(10^6);
```

Consider the following two random variables

```
> X := RandomVariable(Logistic(0, 1));
> Y := RandomVariable(Logistic(10, 2));
```

Let's compare the two densities.



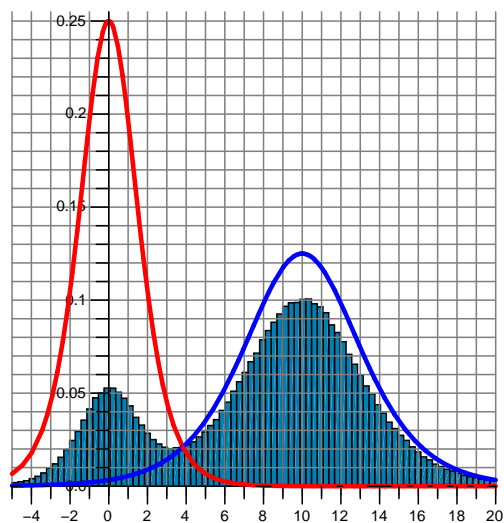
We would like to consider a new random variable, which is X with probability $1/5$ and Y with probability $4/5$.

```
> U := RandomVariable(Bernoulli(1/5)):  
> Z := U*X+(1-U)*Y:
```

We can generate a random sample drawn from this distribution.

```
> S := Sample(Z, ^ (10, 6));
```

Here is the result of the simulation.



REFERENCES

- [1] Karian, Z. A., and Tanis, E. A., *Probability and Statistics: Explorations with MAPLE (2nd ed.)*, 1999, New York: Wiley.
- [2] Rose, C., and Smith, M. D., *Mathematical Statistics and Mathematica*, 2001, New York: Springer-Verlag.
- [3] Tanis, E., *A Review of Maple V, Release 5*, *The American Statistician*, 53, 1999, 389-392.
- [4] Monagan, M., Geddes, K., Heal, M., Labahn, G., Vorkoetter, S., McCarron, J., and DeMarco, P., *Maple 10 Programming Guide*, Maplesoft, 2005.