Cheat sheet Distributions STAT 703

Abraham Arbelaez

Strategies on how to approach questions

1. Identify the **distribution** of the random variable
2. Once you identify the distribution of the random variable in the question, **get familiarized with the distribution** (know whether it is discrete or continuous, how many parameters it should have, examples of the distribution in real life, etc. . .)
3. Identify the **parameters** that are in the question.
4. Read the prompt clear and know what it is asking. We will have two options:

    a. If you have to find the **probability** that X is greater, less or equal to a given number, we will find the probability by using the function that starts with `p`, that is, it could be any of `pbinom()`, `pnorm()`, `pt()`, `pf()`, `pchisq()`, based of our distribution.

    i. If you have to find the **probability** that X is **greater than** a certain number a, this is $P(X > a)$ = area under the curve to the right of a. That is in the general form for coding: `1-pbinom(a,n,p)`, `1-pnorm(a,mean,sd)`, `1-pt(a,df)`, `1-pf(a,df1,df2)`, `1-pchisq(a,df)`

    ii. If you have to find the **probability** that X is **less than** a certain number a, this is $P(X < a)$ = area under the curve to the left of a. That is in the general form for coding: `pbinom(a,n,p)`, `pnorm(a,mean,sd)`, `pt(a,df)`, `pf(a,df1,df2)`, `pchisq(a,df)`

    iii. If you have to find the **probability** that X is **equal to** a certain number a, it means that $P(X = a)$ belongs to a discrete distribution (binomial for this class). So you use `dbinom(a,n,p)`

    iv. If you have to find the **probability** that X is **in between** two numbers a and b, this is $P(a < X < b)$ = area under the curve in between a and b. We will subtract the function with the highest value to the lowest. That is in the general form for coding: `pbinom(b,n,p)-pbinom(a,n,p)`, `pnorm(b,mean,sd)-pnorm(a,mean,sd)`

    b. If you are asked about some **value** a and we know P(X something a) = a given probability p, we will have to find the quantile by using the function that starts with `q`, that is, it could be any of `qnorm(p,mean,sd)`, `qt(p,df)`, `qf(p,df1,df2)`, `qchisq(p,df)`, based of our distribution.

    i. If you are asked about some **value** a and we know P(X > a) = a given probability p, it means that we wnow the area under the curve to the right of x. To calculate a we could use `qnorm(1-p,mean,sd)`, `qt(1-p,df)`, `qf(1-p,df1,df2)`, `qchisq(1-p,df)`

    ii. If you are asked about some **value** a and we know P(X < a) = a given probability p, it means that we wnow the area under the curve to the left of a. To calculate a we could use `qnorm(p,mean,sd)`, `qt(p,df)`, `qf(p,df1,df2)`, `qchisq(p,df)`

Examples following the strategies

- *A random variable T follows a t-distribution with 25 degrees of freedom. What is the value of t in the probability statement* $P(T < t) = 0.95$*?*

1. We first identify the distribution of the random variable. T follows a **t-distribution**.

2. We get familiarized with the distribution. The t-distribution resembles the normal distribution but has heavier tails. It also has one parameter degrees of freedom df.

3. We identify the parameters. Our parameter of interest is **25 degrees of freedom**.

4. Since the probability is given, it is the 4b option. Since $P(T < t)$, we can infer that it is the *ii option*. We now use the function for the t-distribution and plug the values $p = 0.95$ and $df = 25$:

```
qt(0.95, 25)
```

```
## [1] 1.708141
```

Therefore, $t = 1.708$. In other words, the point where the area under the curve equals 0.95 is 1.708

- *The random variable X follows an F distribution with numerator degrees of freedom 5 and denominator degrees of freedom 10. Find the probability that X is less than 0.35, that is* $P(X < 0.35)$

1. We first identify the distribution of the random variable. X follows an **F distribution**.

2. We get familiarized with the distribution. The F-distribution is a probability distribution that has positive numbers only. It has two parameters: the numerator degrees of freedom $df1$, and the denominator degrees of freedom $df2$.

3. We identify the parameters. Our parameters of interest are **df1 = 5, df2 =10**.

4. Since they are asking to find the probability, it is the 4a option. Since $P(X < 0.35)$, we can infer that it is the *ii option*. We now use the function for the F distribution and plug the values $a = 0.35$, $df1 = 5$, and $df2 = 10$:

```
pf(0.35,5,10)
```

```
## [1] 0.1290566
```

Therefore, $P(X < 0.35) = 0.129$. In other words, the area under the curve to the left of the point $a = 0.35$ is 0.129

Distributions

Discrete (whole numbers)

Discrete distributions represent outcomes that are distinct and separate, often associated with countable events or variables. They are characterized by probability mass functions, where probabilities are assigned to individual outcomes or intervals.

Binomial

The binomial distribution models the number of successes in a fixed number of independent trials, where each trial has two possible outcomes (success or failure) with a constant probability of success. It is characterized by two parameters: the number of trials n and the probability of success p.

$$X \sim \text{Binomial}(n, p)$$

Mean : $\mu = np$

Standard deviation: $\sigma = \sqrt{np(1-p)}$

Variance: $\sigma^2 = np(1-p)$

Continuous (real numbers)

Continuous distributions model variables that can take on any value within a specified range, typically associated with measurements. They are described by probability density functions, where probabilities correspond to intervals rather than specific points.

Uniform

In a uniform distribution, all outcomes within a given range are equally likely to occur, resembling a flat, constant probability density across that range. It is characterized by two parameters: the minimum value of the range a and the maximum value of the range b.

$$X \sim \text{Uniform}(a, b)$$

Mean: $\mu = \frac{a+b}{2}$

Standard deviation: $\sigma = \frac{b-a}{\sqrt{12}}$

Variance: $\sigma^2 = \frac{(b-a)^2}{12}$

Normal

The normal distribution, also known as the bell curve, is a symmetrical probability distribution characterized by its mean μ and standard deviation σ.

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

Mean: μ

Standard deviation: σ

Variance: σ^2

Student's t distribution

The Student's t-distribution is a probability distribution used in statistics to estimate population parameters when the sample size is small or the population standard deviation is unknown. It resembles the normal distribution but has heavier tails, making it more robust against outliers and providing more accurate estimates for smaller sample sizes. It has one parameter degrees of freedom df.

$$T \sim \text{t distribution with 'df' degrees of freedom}$$

F distribution

The F-distribution is a probability distribution that has positive numbers only. It has two parameters: the numerator degrees of freedom $df1$, and the denominator degrees of freedom $df2$.

$$F \sim \text{f distribution with 'df1' and 'df2' degrees of freedom}$$

Chi-squared

The Chi-squared distribution is a probability distribution that has positive numbers only. It has one parameter: degrees of freedom df.

$$X \sim \chi^2(\text{'df' degrees of freedom})$$