Expectation and variance

Expected value and variance of common distributions

Geometric distribution

Let X be geometrically distributed with the parameter p. The expectation is then

$$E[X] = \frac{1}{p}$$

where p is the probability of success.

Let X be geometrically distributed with the parameter p. The variance is then

$$Var[X] = \frac{1-p}{p^2}$$

where p is the probability of success.

Exponential distribution

Let X be exponentially distributed with the parameter λ. The expectation is then

$$E[X] = \frac{1}{\lambda}$$

Let X be exponentially distributed with the parameter λ. The variance is then

$$Var[X] = \frac{1}{\lambda^2}$$

Normal distribution

Let X be a normally distributed random variable, or $X \sim N(\mu, \sigma 2)$. The expectation is then

$$E[X] = \mu$$

Let X be a normally distributed random variable, or $X \sim N(\mu, \sigma 2)$. The variance is then

$$Var[X] = \sigma^2$$

Expected value

To find the expected value of a probability distribution, we can use the following formula:

$$\mu \text{ or } E[X] = \sum_i a_i \cdot P(X = a_i) = \sum_i a_i \cdot p(a_i)$$

where a_i is the data value and $P(a_i)$ is the probability of that value.

Example:
Here we have a sample data set of the goal probability of a football team:

Table 1: Goal probability of a football team

Goals (X)	Probability P(X)
0	0.18
1	0.34
2	0.35
3	0.11
4	0.02

Method 1 - using `weighted.mean()`

The first method to calculate the expected value is using the `weighted.mean()` function as shown:

```
#define values
vals <- c(0, 1, 2, 3, 4)

#define probabilities
probs <- c(.18, .34, .35, .11, .02)

#calculate expected value
weighted.mean(vals, probs)
```

```
## [1] 1.45
```

Sample mean

We can calculate the sample mean using the `mean()` function like this:

```
#define values
vals <- c(3, -2, -5, 2, 5, 2, 5, -1, -3, 4, 2)

#calculate sample mean
mean(vals)
```

```
## [1] 1.090909
```

Variance

The variance is a way to measure how spread out data values are around the mean. The formula to find the variance of a population is as follows:

$$\sigma^2 \text{ or } Var(X) = E[X^2] - E[X]^2$$

If the generator of the random variable X is discrete with a probability mass function $(x_1 \to p_1, x_2 - > p_2 ... x_n - > p_n)$ then the variance can be calculated as such:

$$Var(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2$$

Where μ is the expected value i.e.:

$$\mu \text{ or } E[X] = \sum_i a_i \cdot P(X = a_i) = \sum_i a_i \cdot p(a_i)$$

where a_i is the data value and $P(a_i)$ is the probability of that value.

Sample variance

Suppose we have the following dataset:

Table 2: Sample dataset

2	4	4	7	8	12	14	15	19	22

We can then calculate the sample variance using the `var()` function:

```
#define dataset
data <- c(2, 4, 4, 7, 8, 12, 14, 15, 19, 22)

#calculate sample variance
var(data)
```

```
## [1] 46.01111
```

Population variance

Suppose we have the same dataset as previously we can then calculate the population variance by simply multiplying the sample variance by `(n-1)/n`:

```
data <- c(2, 4, 4, 7, 8, 12, 14, 15, 19, 22)

#determine length of data
n <- length(data)

#calculate population variance
var(data) * (n - 1) / n
```

```
## [1] 41.41
```