Forecasting Stats Review Guide

A.Y

July 2018

Random Variables

- Defn: A variable whose value is unknown
- It's the set of possible values that a RV can take is the sample space, X
- Note that an RV can be either discrete or continuous
  - **Discrete:**
    * pmf = probability mass function = $P_X(x) = p(X = x)$
    * A discrete RV has a sample space that is countable, for e.g: $X = \{0, 1\}$ or $X = \{1, 2, 3\}$
  - **Continuous:**
    * pdf = probability density function = $f_X(x) \to p(a < x < b) = \int_a^b f_X(x)\ dx$
    * A continuous RV has a sample space that contains an uncountable number of elements, for e.g: $X = [0, 1]$ or $X = \mathbb{R}$
- An RV, X, can take any value in the sample space, X
- The probability distribution of an RV tells us which values are more likely and which values are less likely
- an event is an outcome or a set of outcomes of an RV
- For a **discrete RV** we can ask what is the probability of the event that the RV, X is = to x. That is, $Pr(X = x)$
- For a **continuous RV** we cannot assign probability to any one event. Thus, we assign probability to a set of events. That is, $Pr(x_1 \leq X \leq x_2)$

The Probability Distribution of a Discrete RV

- For a discrete RV we can assign a probability to each of the countable possible outcomes
- One way to think about this is *relative frequency*. That is, how many times have you observed a particular event out of all the times you've observed the RV
- The probability assigned to each event is summarized by the ***pmf (probability mass function)***
- The pmf is defined as $P_X(x) = Pr(X = x)$
- Any function, $P_X(x)$ is a pmf if:
  - $P_X(x) \geq 0 \forall x \in X$
  - $\sum_{x \in X} P_X(x) = 1$
- These two conditions guarantee that the probability of any event is between 0 and 1 such that the sum of all probabilities is = 1.
  - Note $\sum_{i=1}^{N} x_i = x_1 + x_2 + ... + x_N$
- Using the pmf we can then assign probabilities to events

- For example, let the "event" be that the random var, X, takes a value bigger than x_1 and smaller than x_2. We call this event A, $A = \{x \in X : x_1 \leq x \leq x_2\}$
- Then, the probability of the event, A is: $Pr(A) = \sum_{x \in A} P_X(x)$
- That is, the probability of an event is just the sum of the probabilities of all the possible outcomes that make up the event
  - For example, an event might be: The RV takes a value less than x
  - That is, $A = \{z \in X : z \leq x\}$
  - This event consists of all the possible outcomes of the RV whose value is smaller than x
  - The probability of this event is just the sum of the probabilities of all outcome whose value is smaller than x
- We define the probability of such an event, A, as the ***cumulated probability***that $X \leq x$ and we denote this cumulated probability $F(x)$
- The **cumulative distribution function (CDF)** is the function $F(x) \forall x \in \mathbb{R}$

An Example

Let X be a discrete random variable with sample space $X = \{1, 4, 8, 10\}$

Suppose further that his RV has a probability mass function given by the following:

x	1	4	8	10
$P_x(x)$	0.1	0.2	0.6	0.1

Let A be an event that the RV, X takes a value greater than 3 and less than 10.

We write it as the following:
$$A = \{x \in X : 3 \leq x \leq 10\} = \{4, 8, 10\}$$

Then:
$$Pr(A) = \sum_{x \in A} P_X(x)$$
$$= P_X(4) + P_X(8) + P_X(10)$$
$$= 0.2 + 0.6 + 0.1 = 0.9$$

We construct the CDF for X in a similar way

$\forall x < 1, \{z \in X : z \leq x\}$ is the empty set

Thus, $F(x) = 0 \forall x < 1$

For all values $1 \leq x \leq 4$ the set of outcomes $\{z \in X : z \leq x\} is the set \{1\}$ which has a probability equal to 0.1

Thus, $F(x) = 0.1 \forall 1 \leq x \leq 4$

For all values $4 \leq x \leq 8$ the set of outcomes $\{z \in X : z \leq x\} is the set \{1, 4\}$ which has a probability equal to 0.3

Thus, $F(x) = 0.3 \forall 4 \leq x \leq 8$

For all values $8 \leq x \leq 10$ the set of outcomes $\{z \in X : z \leq x\} is the set \{1, 4, 8\}$ which has a probability equal to 0.9

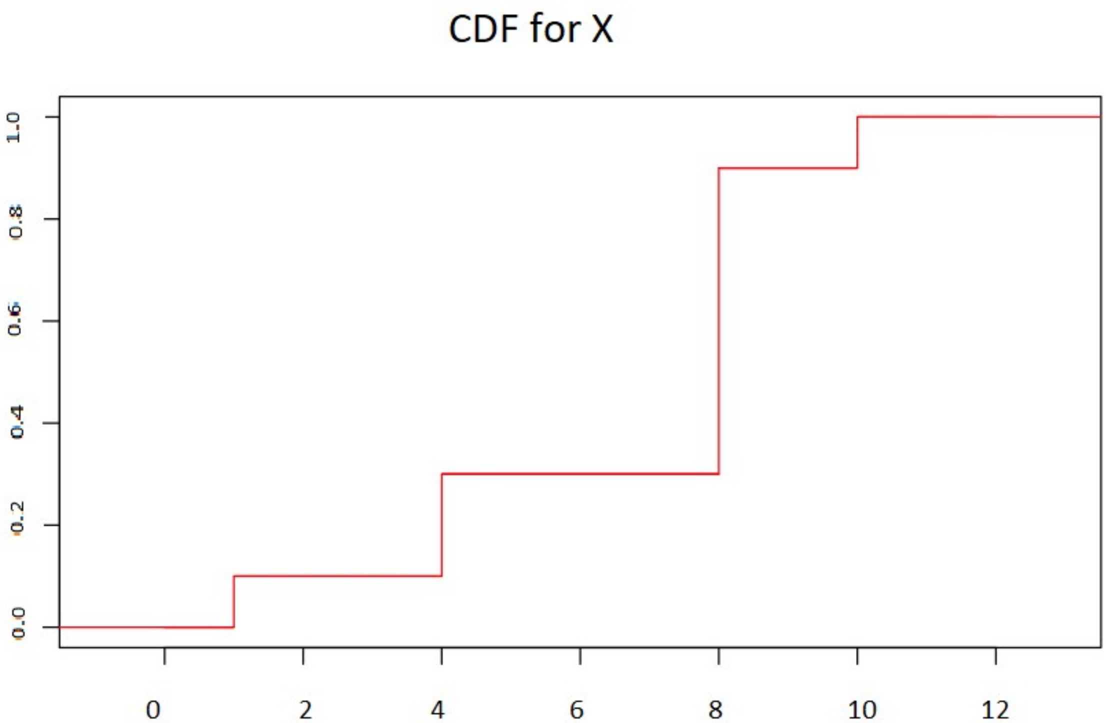
Thus $F(x) = 0.9 \forall 8 \leq x < 10$

Finally, $\forall x \geq 10$ the set of outcomes $\{z \in X : z \leq x\}$ is the set $\{1, 4, 8, 10\}$, which has a probability $= 1$

Thus, $(F(x) = 1)\forall x \geq 10$

So, the CDF for X is: $F_X(x)$

- $0, x < 1$
- $0.1, 1 \leq x < 4$
- $0.3, 8 \leq x < 10$
- $1, x \geq 10$



The Probability Distribution of a Continuous RV

For a continuous RV, there are uncountably many outcomes. Thus, for an event that is comprised of a single outcome, e.g: $A = \{x_0\}$ we have $Pr(A) = 0$

For a continuous rv we can only assign probability to intervals:

- An interval *[a, b]* is the set of real numbers:
- $\{x \in \mathbb{R} : a \leq x \leq b\}$

  For a continuous RV, we describe a pdf, $f_x(x)$ s.t.,:

$$Pr([a, b]) = \int_a^b f_x(z)dz$$

The probability density defines the "increment" to probability of adding x to the interval

As in the discrete case, we can define the cdf as:

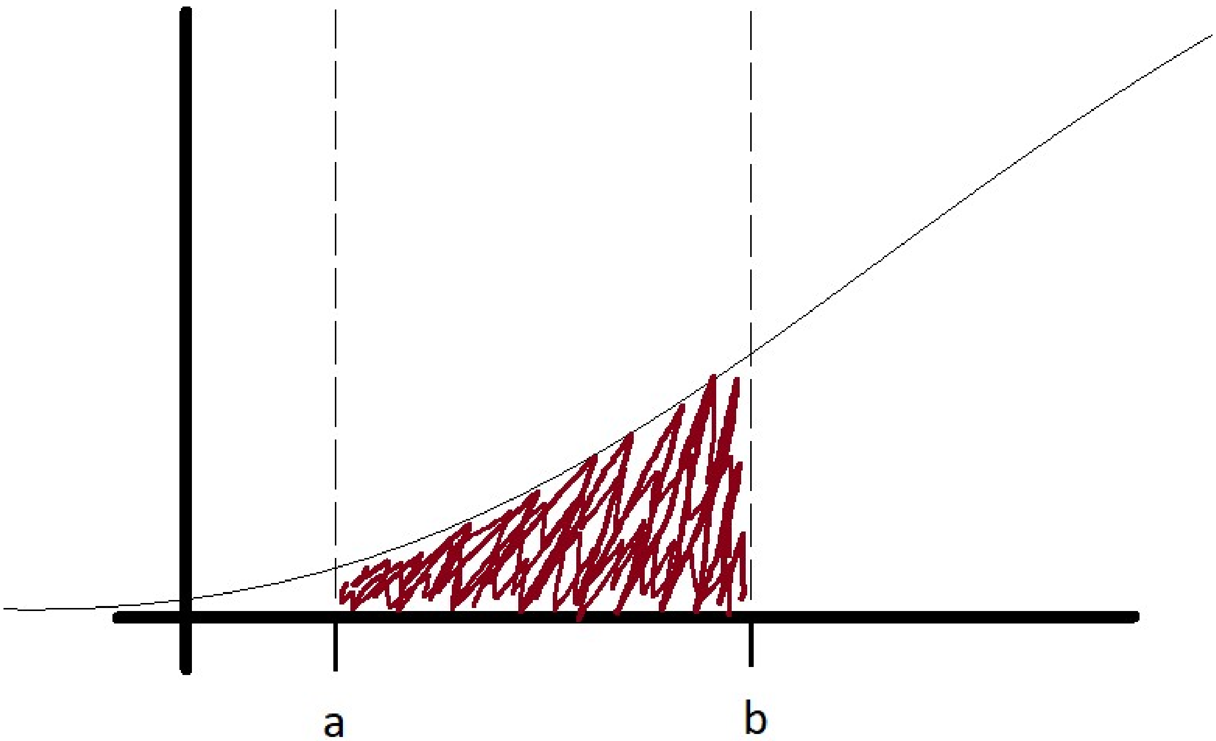
$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz$$

The density(pdf) is therefore just the derivative of the cdf. To build up the probability distribution of a continuous random variable we work with the CDF. Over time we can observe many outcomes and work out the relative frequency of an outcome falling in a particular interval, [a, b]. We can (at least theoretically) do this for any and all such intervals.

It can be shown that the probability of an interval is:

$$Pr([a,b]) = F_X(b) - F_X(a)$$

Thus, any cdf has to be such that the above relationship is satisfied for all intervals.

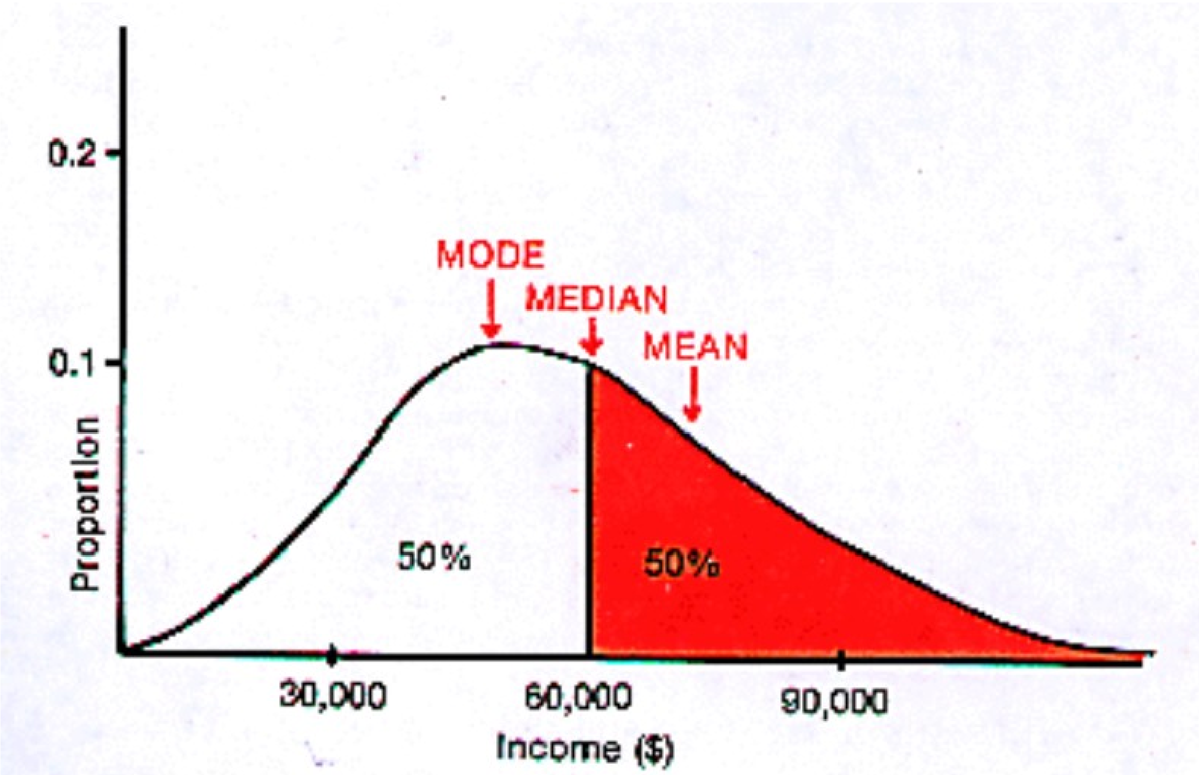


Characterizing the Probability Distribution of the RV

- There are many ways to characterize the probability distribution of a RV
  - plot the pmf/pdf
  - plot the cdf
  - give a mathematical form to the pmf/pdf or CDF
- We can also calculate the "moments" of the distribution
  - Expected Value (first moment)
  - Variance (central 2nd moment)
  - Skewness
  - Kurtosis

Central Tendency of a Distribution

Note that the Central Tendency of a distribution looks like this:



An important characteristic of the distribution of a RV is its central tendency. This can be measured in a number of ways:

- Expected value (mean). This is:

$$\mathbb{E}(X) = \sum_{x \in X} x P_X(x)$$

or

$$\mathbb{E}(X) = \int_x z f_x(z) dz$$

- **Median**: This is the value x_{med} that satisfies $F_X(x_{med}) = 0.5$
- **Mode**: This is the value x_{mode} that maximizes P_X or f_X. It is the outcome that is the most likely

The central location of a distribution is very useful to us when forecasting. What is the best guess if you had to quickly make a forecast? You can use either of the central measures described above - $\mathbb{E}, Median, or Mode$. Forecasts based on these three measures have different properties which we ill look at later in the course. As always, your choice is dependent on the needs of your client.

The Joint Distribution between 2 RVs

Suppose now we have 2 RVs, X and Y

We can examine the probability distributions of X and Y in isolation. These distributions are called marginal distributions.

X and Y may be related and move together in various ways. To investigate this relationship we study the **joint probability distribution of X and Y**. To do so, we must define a joint pdf $F_{XY}(x, y)$, which is:

$F_{XY}(x, y) = Pr(X \leq x \text{ and } Y \leq y$

Based on the jdf in most cases, we can define either a joint density, $f_{XY}(x, y)$ or a joint pmf, $P_{XY}(x, y)$

These functions fully describe the joint distribution between the two RVs X and Y

Measuring the relationship between 2 RVs

If 2 RVs are related we would like a way of measuring this relationship. We can call this measure **covariance**.

The covariance between two RVs is:

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

The sign of this measure tells us whether the 2 RVs are positively/negatively related

Covariance is not invariant to how the variables are measured. Thus, we can't tell anything about the strength of the relationship from the magnitude of the measure (This is correlation).

So we define **correlation**. The correlation between 2 RVs is :

$$\rho xy = corr(X, Y) = \frac{cov(X, Y)}{\sigma_x \sigma_y}$$

Correlation is just the covar divided by the std.dev of the 2 RVs. We can show that no matter the units that:

$$-1 \leq \rho xy \leq 1$$

The closer the correlation is to 1, or -1, the stronger the relationship

Conditional Probability Distribution

We can also define the conditional probability distribution. The conditional density of function of Y conditional on the value of X=x is:

$$f_{Y|X=x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}$$

That is, the conditional probability is just the joint probability divided by the marginal probability of the conditioning variable.

Estimating Moments of a Distribution:

We don't know the tru popl distribution of a RV, no do we know the parameters that govern the distribution. Suppose that we observe a sample from the population, $\{x_1, ..., x_N\}$.

The sample average, $\bar{X}$ is an **estimator** for the population mean (exp. val) where:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

The sample variance, S_X^2 is an estimator for the population variance. That is:

$$S_X^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^2$$

The sample standard deviation, S_X os"

$$S_X = \sqrt{S_X^2}$$

The sample skewness is:

$$S_k = \frac{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^3}{S_X^3}$$

The sample kurtosis is:

$$K = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^4}{S_X^4}$$

The sample Covar is:

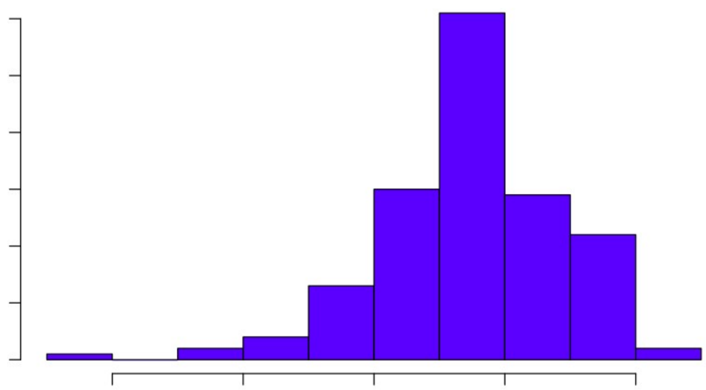
$$S_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})$$

The sample Corr is:

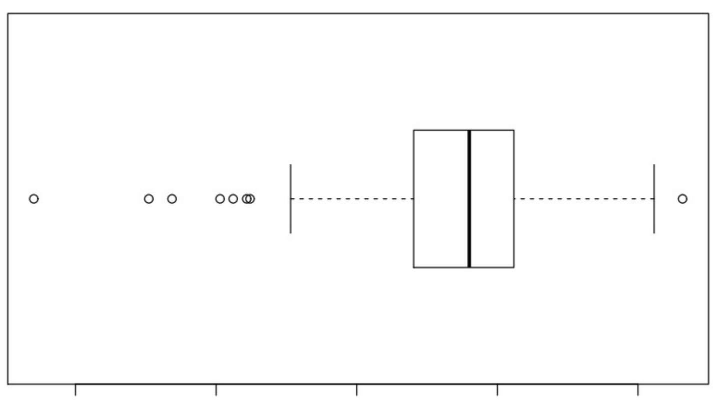
$$\hat{\rho}XY = \frac{S_{XY}}{S_X S_Y}$$

We can also estimate the underlying distribution functions of a RV from a sample. We can estimate and plot the **empirical cdf**. We can plot a histogram or a boxplot. We can also estimate a plot a kernal density estimate:

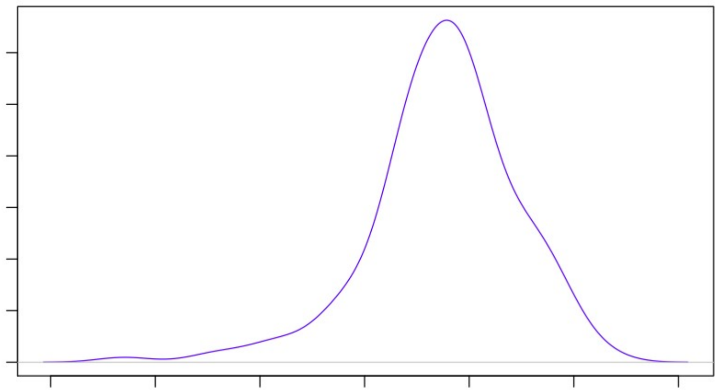
Histogram of US Consumption Data



Boxplot of US Consumption Data



Density Plot of US Consumption Data



Empirical CDF of US Consumption Data