study

Preliminaries (Video, Dr.)

Standard Mathematical and Statistical notation:

Vector and Matrix Notation

- A *scalar* is a number. *Scalars* are represented by lower case letters from the beginning of the alphabet such as a, b, c etc.
- A *vector* is a $n \times 1$ array defined with the mathematical operations of addition and multiplication. The standard convention is for all vectors to be column vectors, i.e. they are 'long' with n rows and 1 column. Vectors are represented as **bold** faced lower case letters frequently from the end of the alphabet, such as $\mathbf{x}, \mathbf{u}$, and $\mathbf{v}$. the ith entry of a vector $\mathbf{u}$ is denoted by $\mathbf{u}[i] = u_i$.
- A *matrix* is a $n \times m$ array defined with the mathematical operations of addition and multiplication. Matricies are represented by a bold face upper cased letter such as $\mathbf{A}, \mathbf{W}, \mathbf{X}$, etc. The (i, j)th entry of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}[i, j] = a_{ij}$.
- The transpose of a $(n \times 1)$ column vector $\mathbf{a}$ is the $(1 \times n)$ row vector $\mathbf{a}^T = [a_1 \dots a_n]$. Sometimes the transpose $\mathbf{a}^T$ is denoted by $\mathbf{a}'$.
- The transpose of a $(n \times m)$ matrix $\mathbf{A}$ is the $(m \times n)$ matrix $\mathbf{A}^T$ where $\mathbf{A}[i, j] = \mathbf{A}^T[j, i]$. When a matrix is transposed, the rows become the columns and the columns become the rows.
- It is preferred to use the T notation $\mathbf{a}^T$ instead of the "prime notation" $\mathbf{a}'$.

Random Variable Notation

Random variables are how you develop calculus based probability theorey.Random variables are the unkown statistical experiement that generate data. Statistical theory is based upon the concept of random sampling.

- Random variables are denoted by capital letters from the end of the alphabet such as U, V, X, Y, or Z.
- The *observed value* of a random variable is denoted by the lower cased counterpart such as u, v, x, y, or z.
- When we have a *random sample* of independent and identically distributed (iid) random variables, we will index the variables in a set such as $X_1, X_2, \dots, X_n$ for the random variables and $x_1, x_2, \dots, x_n$ for the observed values.
- Random variables are used to devleop statistical estimators. Observes values of random variables are used to compute statistical estimates.
- Random variable notation can become convoluded when we move to multivariate random variables. Pay attention to how an author presents these concepts in text.

'Distribution'

The term *distribution* is used throughout all statistical applications and discussions. Loosly speaking, the term distribution is meant to describe how a group of values are related to either each other or to the range of values on which they are defined (their *support*).

The term *distribution* is used rather sloppily. If you don't understand the context you wont understand the use. The term *distribution* is mapped to many related concepts. In general the term *distribution* is related to the characterization of a random variable, or data generated by a random variable.

There are many mathematical notations for characterizing a statistical distribution. The choice of characterization will depend on the context and the existence of the characterization. A random variable can be characterized by any of the following functions.

- The *cumulative distribution function* (cdf), denoted by $F(x) = Pr(X \leq x)$. the cdf will exist for all random variables, and in general is why we use the term "distribution" so looslely throughout statistics. cdf exists for all random variables. From data you can always estimate a distribution function.
- The *probability density function* (pdf) for continuous random variables, denoted by $f(x)$, or the *probability mass function* (pmf) for discrete random variables, denoted by $p(x)$. Note that neither of these functions are guaranteed to exist. A random variable that can be described with a cdf will not always possess a pdf or pmf.
- Transformation functions such as the moment generating function $m(t) = \mathbb{E}[\exp(tX)]$ and the characteristic function $\phi(t) = \mathbb{E}[\exp(itX)]$. Transformation functions are not used when working with data, they may be used to develop conceptual underpinnings of modeling.
- Specialized representations for particular applications such as the *hazard fucntion* $h(t) = \frac{f(t)}{S(t)}$ and the *survival function* $S(t) = 1 - F(t)$ used in Survival Analysis. Survival function is a simple map of the cdf. The hazard function allows you to get a generic representation of a survival function.
- In data analysis distributions can be analyzed using the empirical cdf, the histogram, the Quantile-Quantile plot, and the Kolmogorov-Smirnov test.
- If you need to assess the distribution of residuals in linear regression and compare that to the assumption that they are normally distributed.

Mathematical Expectation

Mathematical Expertation is the theoretical averaging of a random variable with respect to its distribution function. In this sense the pdf of pmf act as a weight function that allows you to find the "center" of the distribution.

For a continuous random variable X with pdf function $f(x)$, the mathematical expectation of X can be computed by

$$\mathbb{E}[X] = \int x f(x) dx$$

For a discrete random variable X with pmf function $p(x) = Pr(X = x)$, the mathematical expectation of X can be computed by

$$\mathbb{E}[X] = \sum_x x p(x)$$

$\mathbb{E}[X]$ is also referred to as the first moment of X.

Expectation, Variance, and Covariance as Mathematical Operators

Let X denote a random variable. Consider the affine transformation $aX + b$.

- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- $\text{Var}[aX + b] = a^2\text{Var}[X]$

Let X and Y be random variables with a joint distribution function. (In the continuous case we would denote this joint distribution function by the join density function $f(x, y)$.) Consider the linera transformtions aX and bY.

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- $\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y] + ab\text{Cov}[X, Y]$

Here the reader should note that in general $\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y]$. If X and Y are independent random variables, then $\text{Cov}[X, Y] = 0$. The converse of this statement is not true except when both X and Y are normally distributed. In general $\text{Cov}[X, Y] = 0$ does not imply that X and Y are indepdented random variables.