Math 461 Notes

Alexander Roe

Data Analysis

Definition. Let X be a data set. The cumulative distribution function, or CDF, is defined as

$$F(x) = P(X \leq x)$$

That is, it measures the probability that a member of X is less than or equal to x. Thus, $1 - F(x)$ measures the probability that a member of X is greater than x.

Fact. The plot of a CDF never decreases, and is bounded by $0 \leq F(x) \leq 1$.

Definition. Let X be a data set with members $x_1, x_2, ..., x_n$ occurring with probabilities $p_1, p_2, ..., p_n$. The expected value of X is defined as

$$E[X] = \sum_{i=1}^{n} x_i p_i$$

That is, the expected value is the arithmetic mean of the members of X.

Definition. Let X be a data set, and let $\mu = E[X]$. Then the variance of X is defined as

$$Var(X) = E[(X - \mu)^2]$$

Informally, variance measures how spread apart the members of X are from the expected value.

Theorem. Let X be a data set, and let c be some real constant. Then, the following properties hold:
(i) $E[cX] = cE[X]$
(ii) $E[X + c] = E[X] + c$
(iii) $Var[X + c] = Var[X]$
(iv) $Var[cX] = c^2 Var[X]$
(v) $Var[X] = E[X^2] - E[X]^2$

Probabilities

More Data Analysis

Simulations

Normal and Exponential

Random Variables

Joint Distributions

Central Limit

Counting

Statistics