

Implementations of Probability Theory

Independent Study Report
Andrew Simonson

Compiled on: October 4, 2024

Contents

	Page
1 Objective	2
2 Units	3
2.1 Unit 1: Statistics Review	3
2.1.1 Random Variables	3
2.1.2 Sample Space	3
2.1.3 Probability Axioms	3
2.1.4 Expectations and Deviation	4
2.1.5 Probability Functions	4
2.1.6 Limit Theorems	5
2.1.7 Confidence	5
2.1.8 Statistical Inference	5

1 Objective

yada yada yah I started this independent study for my own selfish gain

2 Units

2.1 Unit 1: Statistics Review

To ensure a strong statistical foundation for the future learnings in probabilistic models, the first objective was to create a document outlining and defining key topics that are prerequisites for probabilities in statistics or for understanding generic analytical models.

2.1.1 Random Variables

1. **Discrete Random Variables** - values are selected by chance from a countable (including countably infinite) list of distinct values
2. **Continuous Random Variables** - values are selected by chance with an uncountable number of values within its range

2.1.2 Sample Space

A sample space is the set of all possible outcomes of an instance. For a six-sided dice roll event, the die may land with 1 through 6 dots facing upwards, hence:

$$S = [1, 2, 3, 4, 5, 6] \quad \text{where } S \text{ is the sample space}$$

2.1.3 Probability Axioms

There are three probability axioms:

1. **Non-negativity:**

$$P(A) \geq 0 \quad \text{for any event } A, P(A) \in \mathbb{R}$$

No event can be less likely to occur than an impossible event ($P(A) = 0$). $P(A)$ is a real number. Paired with axiom 2 we can also conclude that $P(A) \leq 1$.

2. **Normalization:**

$$P(S) = 1 \quad \text{where } S \text{ is the sample space}$$

Unit Measure - All event probabilities in a sample space add up to 1. In essence, there is a 100% chance that one of the events in the sample space will occur.

3. **Additivity:**

$$P(A \cup B) = P(A) + P(B) \quad \text{if } A \cap B = \emptyset$$

A union between events that are mutually exclusive (events that cannot both happen for an instance) has a probability that is the sum of the associated event probabilities.

2.1.4 Expectations and Deviation

1. **Expectation** - The weighted average of the probabilities in the sample space

$$\sum^S P(A) * A = E \quad \text{where } E \text{ is the expected value}$$

2. **Variance** - The spread of possible values for a random variable, calculated as:

$$\sigma^2 = \frac{\sum(X - \mu)^2}{N}$$

Where N is the population size, μ is the population average, and X is each value in the population.

For samples, variance is calculated with **Bessel's Correction**, which increases the variance to avoid overfitting the sample:

$$s^2 = \frac{\sum(X - \bar{x})^2}{n - 1}$$

3. **Standard Deviation** - The square root of the variance, giving a measure of the average distance of each data point from the mean in the same units as the data.

$$\sigma = \sqrt{V} \quad \text{where variance is } V$$

2.1.5 Probability Functions

Probability Functions map the likelihood of random variables to be a specific value.

Probability Mass Functions

Probability Mass Functions (PMFs) map discrete random variables. For example, a six-sided die roll creates a uniform random PMF:

$$P(A) = \begin{cases} 1/6 & \text{if } X = 1 \\ 1/6 & X = 2 \\ 1/6 & X = 3 \\ 1/6 & X = 4 \\ 1/6 & X = 5 \\ 1/6 & X = 6 \end{cases}$$

Probability Density Functions

Probability Density Functions (PDFs) map continuous random variables. For example, this is a PDF where things happen.

$$P(A) = \begin{cases} X & \text{if } 0 \leq X \leq .5 \\ -X + 1 & .5 < X \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

2.1.6 Limit Theorems

Law of Large Numbers

The Law of Large Numbers states that as the number of independent random samples increases, the average of the samples' means will approach the true mean of the population.

$$\text{true average} \approx \frac{1}{n} \sum_{i=1}^n X_i \quad \text{as } n \rightarrow \infty$$

Central Limit Theorem

The Central Limit Theorem states that the sampling distribution of a sample mean is a normal distribution even when the population distribution is not normal.

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

Where X_i is the sample mean, $N(0, 1)$ is a standard normal distribution, and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

This is a challenging to understand solely as an equation. As an example, take a sample of two six-sided dice rolls and average their numbers. The more sample averages taken, the more they will resemble a normal distribution where the majority of samples average around 3.

2.1.7 Confidence

Confidence is described using a confidence interval, which is a range of values that the true value is expected to be in, and its associated confidence level, which is a probability (expressed as a percentage) that the true value is in the confidence interval.

It is important to note that confidence levels, such as 95%, do not indicate that the real value is within 5% of the point estimate. The confidence level expresses the probability that the real value is in the range provided by the confidence interval.

At the highest level, calculating confidence intervals is simply the observed statistic (generally the mean) plus or minus the standard error.

To calculate standard error, kys.

2.1.8 Statistical Inference

Statistical Inference is any data analysis to draw conclusions from a sample to make assertions about the population. Methods include estimation via averages and confidence intervals, and hypothesis testing, which attempts to invalidate (never *validate*) a hypothesis.