CHAPTER 1 Random Variables

A random variable X is a random, quantifiable answer that answers some probabilistic question, like "how many cars will pass here in the next 10 minutes?". The RV will have some underlying theoretical PDF $f(x)$, which will collapse into a single value when observed. Consider it a quantum particle in a superposition. The wavefunction will collapse to a single value when observed, but is before that governed by a theoretical probability distribution.

$$\underbrace{f(x) = N(x;\, \mu = 5,\, \sigma = 1.5) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}}_{\text{Theoretical distribution of RV}} \quad \Rightarrow \quad \underbrace{X_i = [5.345,\, 7.955,\, 3.895,\, 1.065,\, 4.701,\, 1.696, ...]}_{\text{Observed sample of RV}}$$

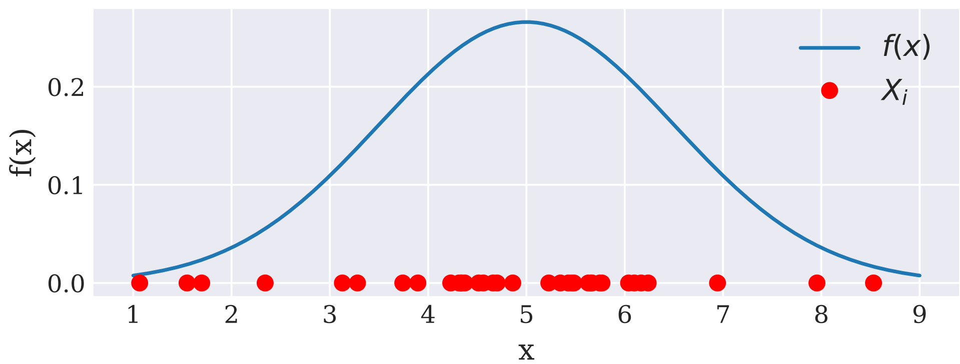

Figure 1.1: Random variable X, with theoretical PDF $f(x)$, and 35 observations, X_i.

Statistics

The measured sample X_i will have statistics such as mean and standard deviation, mirroring the "actual" parameters of the theoretical PDF. They will approach the theoretical values as the sample size increases ($N \gg 1$). The most used statistics are the mean $\overline{X}$ and variance S^2.

Properties of PDF - $f(x)$			**Statistics of RV - X_i**	
Mean	μ	$\leftrightarrow$	$\overline{X}$	Sample Mean
Variance	V	$\leftrightarrow$	S^2	Sample Variance
St.Div.	σ	$\leftrightarrow$	S	Sample St.Div.

Table 1.1: Properties of the RVs PDF, and the corresponding statistics of a sample of the RV.

Distribution of sample statistics

Since each random sample from the RV will be different, the statistics will differ each time as well, as seen in figure 1.2.

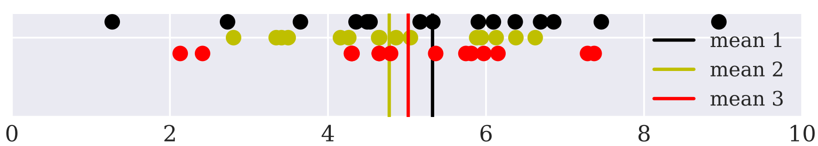

Figure 1.2: Three size 15 sample distributions of random variable X_i

Imagine picking an infitine amount of such random samples from the RV, each of size n. The statistics themselves will now have a distributions, with it's own mean, variance, etc.

Sample mean and Central Limit Theorem

The sample mean of the RV, independently of what sort of distribution the random samples are from, follow

$$E(\overline{X}) = \mu_{\overline{X}} = \mu \qquad V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n \tag{1.1}$$

Sample mean of normally distributed RV

When the RV has a **normal distribution**, the sample mean $\overline{X}$ is **itself normally distributed**, with mean and variance as in 1.1:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

A problem here is that σ is often unknown, as it belongs to the theoretical PDF, not the sample. The **t distribution** solves the problem. The following random variable has a t-distribution when $\overline{X}$ is the sample mean of a normal RV:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Sample mean of any RV - The Central Limit Theorem (CLT)

When the RV has **any distribution**, the sample mean $\overline{X}$ will **approach a normal distribution** for large n. This is called the central limit theorem. We can then use both the Z and T distributions as above.

CHAPTER 2 Point Estimators

Estimators $\hat{\theta}$ are attempts at reconstructing a theoretical parameter (left side of 1.1 - μ, σ,...) from the sample statistics (right side of 1.1 - $\overline{X}$, S,...). The most obvious estimators are simply the mirroring statistics, but there are more advanced ones. Some definitions:

- **Mean Square Error:** $\text{MSE} = E\Big[(\hat{\theta} - \theta)^2\Big] = V(\hat{\theta}) + E(\hat{\theta}) - \theta = \text{variance of estimator} + \text{bias}^2$
- **Unbiased Estimator:** Estimator which variance is zero $V(\hat{\theta}) = 0$.
- **Minimum Variance Unbiased Estimator:** Estimator with lowest variance, which has no bias $E(\hat{\theta}) = \theta$.

Property	**MVUE of norm. dist.**	Other Estimators
Mean μ	$\overline{X}$	$\tilde{X}$, $\overline{X}_{tr(10)}$
Variance σ^2	S^2	
St.Div. σ	–	

Table 2.1: Most common estimators of common properties

Bootstrap point estimation

Moment estimators

Maximum likelihood estimators

CHAPTER 3 Confidence Intervals

The next chapters build on the fact that we now know what sort of distributions our estimators/statistics have.

We can then establish some interval around the estimators' mean where we are $100(1-\alpha)\%$ sure that the actual value lies.

Deriving Confidence Intervals

1. Find some RV which involves only your estimator and known constants, and has a known probability distribution. Ex, for the estimator $\hat{\mu} = \overline{X}$ we can use the RVs

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad \text{or} \qquad T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

   which has a Z and T distributions, depending on if σ is known or not (or N is large).
2. Use the known distributions' critical values to establish an interval. Ex; for the normal case:

$$P\left(z_{1-\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 100(1-\alpha)\%$$

3. Solve the inequality for the value to be estimated, in this case μ:

$$\overline{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad \Rightarrow \qquad \mu \in \left[\overline{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Useful RVs for finding for finding CIs

Distribution	**Random Variable**	Estimator	Comments
Normal Dist.	$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$\hat{\mu} = \overline{X}$	Requires that σ is known.
T Dist.	$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$	$\hat{\mu} = \overline{X}$	$n-1$ df. Approaches Z as $N \gg 1$.
Chi-squared Dist.	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$	$\hat{\sigma^2} = S^2$	$n-1$ df. Take sqrt to get CI for σ.

Table 3.1: RV distributions to use in confidence

Parametric Bootstrap CI

Non-parametric Bootstrap CI

CHAPTER 4 Hypothesis Testing

Introduction

Hypotethis testing is just a formalization of confidence intervals, where we decide if we should throw away some old theory for a new one, because oberved data falls outside a large confidence interval around the old theory.

We have some "previously accepted" **null hypothesis** H_0 about the value of a parameter θ, called the **null value**:

$$H_0 : \theta = \theta_0$$

We present an **alternative hypothesis** H_a, which claims that the null hypothesis is too low, too high, or either:

$$H_a : \theta > \theta_0 \qquad H_a : \theta < \theta_0 \qquad H_a : \theta \neq \theta_0$$

We will have to choose whether to reject or not reject the null hypothesis in favor of the alternative hypothesis. When doing so, one of two errors may occur:

- **Type I Error:** We reject the null hypothesis H_0 when it is true, with probability $\boldsymbol{\alpha}$ of happening.
- **Type II Error:** We do not reject the null hypothesis H_0 when it is false, with probability $\boldsymbol{\beta}$ of happening.

Type I errors are considered the most serious, as it replaces previously accepted knowledge with something new. This means α is the most interesting parameter, and we usually keep it low $(0.01 - 0.1)$.

Rejection Region

We establish $100(1-\alpha)\%$ CI around θ_0, under the assumption that H_0 is true. The area outside this now becomes a "rejection region", where the measured values contradict H_0 enough to reject it.

A rejection region is simply an inverted confidence interval, where we are $(1-\alpha)\%$ sure that we can reject the null hypothesis. We choose a probability α of a type I error that we find acceptable, and thereafter have to reject the null hypothesis if our measurement falls in the rejection region.

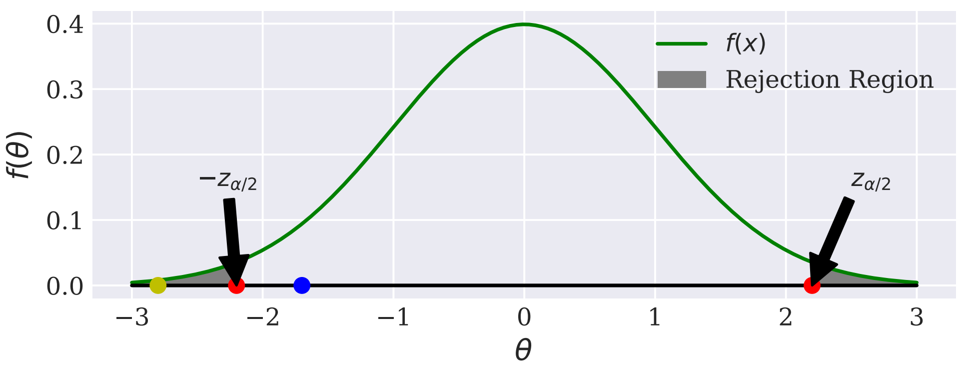

Figure 4.1: Rejection region of normal standard distribution. Yellow and blue dots are examples of two measured values, outside and inside the rejection region.

The distribution is build upon the assumption that H_0 is true, meaning we use the null value $\theta = \theta_0$ in the distribution, and derivation of the confidence interval.

P-Values

The P-value of an alternative hypothesis, is the probability of obtaining a value of the test statistic at least as contradictory to H_0 as the value calculated from the sample.

In simpler terms, it is the area of the rejection region when we place our sample *just* at the edge of the rejection region - $\theta_a = z_{\alpha/2}$.