|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Distribution | pmfpdf | Mean | Variance | Moment Generating Function: ; | Etc. |
|  |  |  |  |  | binary event w/ |
|  |  |  |  |  | of successes in trials. |
|  |  |  |  |  | of events occurring in a fixed time-period with a known rate *.* |
|  |  |  |  |  |  |
|  |  |  |  |  | Memoryless Distribution: |
|  |  |  |  |  | ; |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| CHI-SQUARE DIST.  degrees of freedom |  |  |  |  | &  .  If , then and . |
| t-DISTRIBUTION |  |  |  |  | and : |

|  |  |  |
| --- | --- | --- |
| Prior : | Likelihood : | Posterior : |
|  |  |  |
|  |  |  |
|  |  |  |
|  | is known. | ; |

**Definition**  : . Let be a loss function. For each possible value of let be a value of such that is minimized. Then is called a *Bayes estimator* of . Once is observed is called a *Bayes estimate* of .

**Definition**  : . The loss function is called the *squared loss error*.

**Corollary**  : Let be a real-valued parameter. Suppose that the squared error loss function is used and the posterior mean of , , is finite. Then the Bayes estimator of is .

**Definition**  : . The loss function is called the *absolute error loss*.

For every observed value , the Bayes estimate will now be the value of for which the expectation is a minimum. This is when is the median of the posterior distribution.

|  |
| --- |
| **Proof** : Using the property . Note that in the likelihood proportionality we omit terms only dependent on , i.e., . Hence, . . . Now we complete the squares again. . Since the final term no relation, drop in proportionality. . |

|  |
| --- |
| **Question 1** : Given a random sample . Prior . Let . Find the PMF of . Calculate the posterior distribution and find the Bayes estimator under squared error loss.  **Answer 1** : . . Therefore, . Bayes estimator under square error loss is . Let , the PMF: . . .  **Question 2** : Let and is known. Prior . Find the bayes estimator for under the squared error loss function. **Answer 2** : . Therefore a increases increases if |

|  |
| --- |
| **Question 1** : Show that the distribution is a conjugate prior for iid data from a distribution with unknown parameter . Explicitly state the parameters of the posterior distribution. **Answer 1** : . . . Therefore, . **Question 2** : With the above setup from question 1, find the Bayes estimator of under squared error loss. Next, express as a weighted average of the sample mean and the prior mean, i.e., where is the mean of the prior. What happens to as ? . **Answer 2** : . Therefore, for and , . Thus, by LLN. |

|  |
| --- |
| **Theorem**  : Suppose that the random variables form a random sample from a distribution for which the pdf is . Suppose also that the value of the parameter is unknown and the prior pdf of is . Then the posterior pdf is given by: . We can depict this equivalently by: .Where the proportionality symbol is used to convey that the left-hand side is equal to the right-hand side except possibly up to a constant. The appropriate constant can be determined by using: or . |

|  |
| --- |
| **Confidence Intervals** : meaning = range of values that you expect your estimate to fall between a certain percentage of the time if you run your experiment again.  . If is known and we want a CI for : . as increases thus . If is unknown use : CI for : . where are unknown. Construct CI for . . Since . After observing , this yields a CI for : . |

|  |
| --- |
| **Fisher’s Information**: **Proof** : . Since , then . Therefore, so . Next, note that . So, . As . More information when is larger…  **Theorem** : , i.e., fishers’ information for a random sample is simple times the fishers’ information in a single observation. .  **Theorem** (CRLB) : (iid sample size ) If is an unbiased estimator of and then is a minimum variance unbiased estimator. (BIASED ESTIMATOR) .  **Proof** : Using Cauchy-Swartz inequality, . Set and . We know that the expectation of the score is , i.e., . Thus, . Where (proved above). Therefore, . . Hence, we have proved .WLOG if we have iid sample . prof notes: .  **Theorem** (Asymptotic Distribution of MLE): Suppose the MLE exists and is twice differentiable. Then the asymptotic distribution of . Equivalently, the distribution of . |

|  |  |
| --- | --- |
| **Factorization Theorem** (sufficiency/joint sufficiency) : where for  **parameter exponential family** *:*  **Invariance Property of MLE** ( is one-to-one) : If is the MLE of , then for any function , the MLE of is , e.g. let . . If , then the MLE for is .  **Method of Moments**: Let be a random sample from a population. Method of moment estimation (MOME): Equate sample moments to population moments. If the population has parameters, the MOME consists of solving the system of equations , for the 𝑟 parameters, where , i.e., (kth sample moment) (kth population moment). | . . . It is customary to refer to the rejection region for as the critical region of a test. The probability of obtaining a value of the test statistic inside the critical region when is true is called the **size** of the critical region. Thus, the size of the critical region is just the probability of committing a type I error.  The only way we can reduce the probabilities of both types of errors is to increase the size of the sample, but as long as is held fixed, this inverse relationship between the probabilities of type I and type II errors is typical of statistical decision procedures. In other words, if the probability of one type of error is reduced, that of the other type of error is increased. Note: p-value is area in direction of extremeness of test statistic in direction of alternative calculated assuming is true. If it is two sides we take sum of both directions, i.e., assuming symmetry here. |

|  |
| --- |
| **Theorem** : vs (simple v simple) & iid sample, . Reject if is the most powerful test of size . Note, NP Lemma can be used to test from two different distributions given they are completely specified, i.e., vs .  **Example** (Discrete case) : If is used to test vs of size . Using NP lemma, and thus . To determine a test of size in discrete case, & & & . Therefore, if we want some test of maximum we select such that this exists.  **Theorem** : Suppose we wish to test vs . ; . The statistic is called the likelihood ratio statistic. A likelihood ratio test is to reject if for some constant .  where and . **Prof Notes** : (i) Test vs reject if . (ii) , where  **Example** : Given a random sample and is known. Determine a test of size for vs . **Answer** : and . Some algebra yields . Reject if gives the test reject if . . Therefore, . The test is to reject if ;CI from the compliment, i.e. acceptance region.  **Theorem** : For large , the distribution of converges to the chi-square distribution with degree of freedom, i.e., . For parameters , .  **Example** : Check using for the normal LRT above. |

|  |
| --- |
| **Definition** (Monotone Likelihood Ratio) : Let denote the joint pdf of the observations . Let be a statistic. It is said that the joint distribution of has a *monotone likelihood ratio* (MLR) in the statistic if the following property is satisfied: For every two values and , with , the ratio depends on the vector only through the function , and this ratio is a monotone function of over the range of possible values of . Specifically, if the ratio is increasing, we say the distribution of has increasing MLR, and if the ratio is decreasing, we say that the distribution has decreasing MLR. |

|  |
| --- |
| **Two sample t-test** (equal var): . s.t. .Reject if . Alternatively, use two shorthand : where .  **Proof** : & . **F test** : vs or or . . Reject if ( for both) |

|  |
| --- |
| **Test (** / Use for Test of Independence / Use for testing difference in r proportions between c classes: Let be the probability of the outcome from the population. We want to test for against are not equal for at least one value of . Notation : the observed frequency for the row and column is , the row total is , the column totals is , and the grand total (sum of all cell frequencies) is . With this notation, we estimate the probabilities and as and and under the null hypothesis of independence we get for the expected frequency for the cell in the row and column. The test statistic is given by where we reject if .  **Goodness of Fit Test** : For testing the population follows the specified distribution vs : the population does not follow the specified distribution. The critical region is given by where , is the number of terms in the summation, and is the number of independent parameters estimated on the basis of the sample data. **Example (1)** : test that the data follows . The categories are , , , with observed values , , , respectively. To find the expected counts we find probability of each bin times number of samples, i.e., repeat and calculate , … **Example (2)** : Consider Mendel’s hypothesis that when crossing two types of peas, the probability of the classifications (a) round and yellow, (b) wrinkled and yellow, (c) round and green, (d) wrinkled and green are , , , respectively. If from independent observations, the observed frequemcies of these respective classifications are , , , and , are these data consistent with the Mendelian theory? That is, test with , the hypothesis that the respective probabilities are , , , . **Answer (2)** : For are , , , , with . p-value . Fail to reject , the respective probabilities are , , , . |

|  |
| --- |
| **Simulations** : Simulation is a computational technique that uses models to replicate the behavior of systems, enabling analysis and decision-making when analytical solutions are impractical. It can handle both deterministic systems (fully defined by inputs) and stochastic systems (influenced by randomness). By leveraging the **Law of Large Numbers**, simulation generates and averages random variables to approximate means, distributions, or probabilities. Key features include its ability to capture dynamic and random behaviors and study processes that are complex or impossible to observe directly. Simulation is particularly useful for estimating uncertainty by analyzing the **simulation variance**—the variability in repeated simulations—and the **simulation standard error**, which quantifies the precision of estimates. This makes simulation an essential tool for understanding complex systems and assessing outcomes under uncertainty.  **ANOVA & GLM** : Analysis of Variance (ANOVA) and Generalized Linear Models (GLM) are statistical tools for analyzing relationships between variables. ANOVA tests for significant differences in means across groups by partitioning total variability into components explained by the groups and residual (unexplained) variation. It is commonly used in experimental designs to evaluate the effects of categorical factors. GLM extends linear regression by accommodating various distributions (e.g., normal, binomial, or Poisson) for the response variable, allowing for a broader range of modeling scenarios. GLM uses a link function to relate the predictors to the expected value of the response, enabling flexibility in handling non-normal data. Both techniques assess the significance of predictors and interactions, making them powerful tools for exploring and interpreting data in diverse fields.  **Bootstrapping** is a resampling technique in statistics used to estimate the sampling distribution of a statistic by repeatedly drawing samples with replacement from the original data. It is particularly useful when the theoretical distribution of the statistic is complex or unknown. By generating a large number of bootstrap samples, we can compute estimates of standard errors, confidence intervals, and other measures of uncertainty for parameters such as means, medians, or regression coefficients. Bootstrapping is non-parametric, requiring minimal assumptions about the underlying distribution of the data, and is widely applied in cases where sample sizes are small or traditional methods are impractical. This flexibility makes it a robust tool for statistical inference and model validation.  **The F-test** is a statistical test used to compare the variances of two or more groups or to assess the overall significance of a regression model. In the context of comparing variances, it tests the null hypothesis that the variances of the groups are equal by calculating the ratio of the variances. In the context of regression, the F-test evaluates whether at least one predictor variable significantly contributes to explaining the variability in the response variable, comparing the fit of a model with and without the predictors. The test statistic follows an F-distribution, and a large F-value indicates that the model or group differences are statistically significant. It is commonly used in ANOVA, regression analysis, and comparing nested models.  **Non-parametric** tests are statistical methods used when data doesn't meet the assumptions required for parametric tests, such as normality or homogeneity of variance. These tests are distribution-free, meaning they don’t assume a specific underlying distribution for the data. Common non-parametric tests include the **Mann-Whitney U test** and **Wilcoxon signed-rank test** (for comparing two independent or paired samples, respectively), the **Kruskal-Wallis test** (for comparing more than two independent groups), and the **Spearman's rank correlation** (for assessing the relationship between two variables). Non-parametric tests are often used with ordinal data or when sample sizes are small or skewed. They are robust and flexible, providing valid results even when data is not normally distributed or when outliers are present. |

|  |
| --- |
| **Example (1)**: *Suppose where are unknown.* *Prove and are jointly sufficient.* ANSWER : Since we can equivalently prove and are jointly sufficient as is just a function of and . Therefore, . Simplifies to . Jointly sufficient .  **Example (2)**: Show . . ; . |

|  |
| --- |
| **JACOBIAN TRANSFORMATION** : Given , for which the joint pdf is . Define as and where we assume the functions and are one-to-one. Let the inverse of this transformation be given by and . Then the joint pdf of , is where is the determinant: and denotes the absolute value of the determinant . Thus, the joint pdf is obtained by starting with the joint pdf replacing each value by its expression in terms of , and then multiplying the result by . is called the Jacobian of the transformation. |

|  |
| --- |
| **Theorem** : has the same pdf as the ratio of: divided by where and and are independent . **Proof** : .  **t-distribution** : If given ; and , then (t-distribution with degrees of freedom). This result implies that the pdf is given by (*see chart*) .  **Proof** : . Therefore, the Jacobian is given by . The pdf of and is given by the product of their marginals as . Therefore, where . Therefore, . As a result, plugging in all values: . Thus, which simplifies to . Now we want to find the pdf of so we integrate out the values . . Let and . ; we see this is the which sums to . Therefore, . Plugging in , , yields: . Finally, . Proved . |