CSC591: Foundations of Data Science
HW2: Probability distributions, Expectation, Maximum Likelihood Estimation, Sampling Distribution, Central Limit Theorem, Confidence Intervals, Hypothesis Testing.

Released: 9/23/15
Due: **10/4/15 (23:55pm);** (One day late: -25%; -100% after that).
Note: (R-project (code) can be submitted after mid-term, but by 10/11/15: 23.55pm)

Student Name: Parth Satra
Student ID: 200062999 (pasatra)

Notes

- Filename: Lastname_StudentID.pdf (only pdf).
- You can also submit scanned hand written solution (should be legible, TA's interpretation is final).
- This h/w is worth 5% of total grade.
- You can discuss with your friends, but solution should be yours.
- Any kind of copying will result in 0 grade (minimum penalty), serious cases will be referred to appropriate authority.
- All submission must be through Moodle (you can email to TA with cc to Instructor – only if these is a problem – if not received on time, then standard late submission rules apply)
- No makeups or bonus; for regarding policies, refer to syllabus and 1st day lecture slides.

Q#	Max Points	Your Score
1	5	
2	20	
3	10	
4	10	
5	15	
6	10	
7	10	

Q1. Simple statistics (**5 points**)

(a) List formulae for sample mean, mode, variance, standard deviation. (4 points).
(b) Is the sample variance is an unbiased estimator of population variance? (1 point).

Answer

(a) The image below shows the answer to this part.

Answer 1 (a)

Let n be the total number of elements in the sample &
x represent each element of the sample.

(i) Mean ($\bar{x}$)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

(ii) Mode: Element with highest frequency is called mode. In normal distribution mode is the same as mean and has same above formula.

(iii) Variance (S^2)

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (\bar{x} - x_i)^2$$ where $\bar{x}$ = mean.

(iv) Standard deviation (S)

$$S = \sqrt{S^2} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (\bar{x} - x_i)^2}$$

(b) **YES,** the sample variance is an unbiased estimator of population variance.

Q2. (Expected Values) **(20 points)**

Remember the following (i) and (ii):

(i). If X and Y are two random variables with finite expected values, then E(X+Y) = E(X) + E(Y).

(ii) If X and Y are independent, the E(XY) = E(X)E(Y).

Answer (a) – (h).

- **(a)** Define Expected Value of discrete (numerical) random variable. **(1 point**
- **(b)** Suppose in an experiment a fair coin is tossed 3 times. Let X denotes the number of heads that appeared in the experiment. Then what is E(X). **(2 points)**
- **(c)** Recall the discussion on Bernoulli distribution (W2-C2 lecture). Let S_n be the number of success in n Bernoulli trails with probability p for success on each trail. Then what is $E(S_n)$. **(3 points)**
- **(d)** A coin is tossed twice. Let $X_i = 1$ if the i^{th} toss is heads and 0 otherwise. Then what is $E(X_1X_2)$? **(2 points)**
- **(e)** Let X be a random variable with expected value $\mu = E(X)$. Then what is the Variance of X in terms of E(). **(1 point)**
- **(f)** Let X be a random variable with expected value $\mu = E(X)$, then show that the Variance, $V(X) = E(X^2) - \mu^2$. **(2 points)**
- **(g)** Using the formulae in (f), compute the variance of the outcome of a roll of a die. **(3 points)**
- **(h)** Let X be an exponentially distributed r.v. with parameter λ. Then the density function of X is given by: $f_X(x) = \lambda e^{-\lambda x}$. Compute E(X) and V(X), where V stands for variance. **(3 + 3 = 6 points)**

Answer

Following images give the answer to all parts of question 2.

Answer 2

(a) Let $f(x)$ be discrete probability distribution function, then

$$E(x) = \sum x f(x) \quad \forall x$$

(b) $E(x)$ can be defined as above

$$\therefore E(x) = \sum x f(x) \quad \forall x$$

Here $x = 0, 1, 2, 3$ &

$f(0) = \frac{1}{8}$ $\quad f(1) = \frac{3}{8}$

$f(2) = \frac{3}{8}$, $f(3) = \frac{1}{8}$

$$\therefore E(x) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8}$$

$$= \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = \underline{\underline{1.5}}$$

$$\therefore \boxed{E(x) = 1.5}$$

Answer 2

(c) A bernoulli's trial repeated n times becomes a binomial distribution

$$\therefore P(S_n) = P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Here k is number of successful outcomes after n iterations.

$$\therefore E(S_n) = \sum_{k=0}^{n} k \cdot {}^nC_k \, p^k (1-p)^{n-k}$$

Now at $k=0$ the term will be zero.

$$\therefore E(S_n) = \sum_{k=1}^{n} \frac{k \cdot n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n \cdot (n-1)!}{(k-1)!(n-k)!} p \cdot p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

Let $a = k-1$ & $b = n-1$

$$\therefore \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = \sum_{a=0}^{b} \frac{b!}{a!(b-a)!} p^a (1-p)^{b-a}$$

Thus by Binomial theorem we get

$$\sum_{a=0}^{b} \frac{b!}{a!(b-a)!} p^a (1-p)^{b-a} = (p+1-p)^b = 1^b = 1.$$

$$\therefore \boxed{E(S_n) = np \times 1 = np}$$

Answer 2

(d) Given $X_i = 1$ if i^{th} toss is heads else
$X_i = 0$

$\therefore X_1 = \begin{cases} 1 ; & H \\ 0 ; & T \end{cases}$

$\therefore X_2 = \begin{cases} 1 ; & HH, TH \\ 0 : & HT, TT \end{cases}$

$\therefore E(X_1) = 1 \times 0.5 + 0 \times 0.5 = 0.5$

$\therefore E(X_2) = 1 \times \left(\frac{1}{4} + \frac{1}{4}\right) + 0 \times \left(\frac{1}{4} + \frac{1}{4}\right) = 0.5$

Now $E(X_1 X_2) = E(X_1) \cdot E(X_2)$ since X_1 & X_2 are independent

$$\therefore \boxed{\begin{aligned} E(X_1, X_2) &= 0.5 \times 0.5 \\ &= 0.25 \end{aligned}}$$

(e) Variance of a random variable X can be defined as

$$Var(X) = E\left[(X - E[X])^2\right].$$

Answer 2

(f) Variance of a random variable can be defined as

$$Var(X) = E\left[(X - E[X])^2\right]$$
$$= E\left[(X^2 - 2X E[X] + (E[X])^2\right]$$
$$= E[X^2] - 2E\left[X E[X]\right] + E\left[(E[X])^2\right]$$

Using the formula $E(X+Y) = E(X) + E(Y)$

Now since $(E[X])^2$ is constant

$\therefore E[(E[X])^2] = (E[X])^2$

Thus we get

$$Var(X) = E[X^2] - 2E\left[X E[X]\right] + (E[X])^2$$

Now using $E(XY) = E(X).E(Y)$ we get

$$Var(X) = E[X^2] - 2(E(X).E[E[X]]) + (E[X])^2$$

Since $E(X)$ is constant

$\therefore E[E(x)] = E(X)$

$$\therefore Var(X) = E[X^2] - 2(E(X).E(X)) + (E[X])^2$$
$$= E[X^2] - 2(E[X])^2 + (E(X))^2$$
$$= E[X^2] - (E[X])^2.$$

$$\therefore \boxed{Var(X) = E[X^2] - \mu^2}$$

Answer 2

(g) Now

$$E(X) = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6}$$

$$= 3.5 = 21/6$$

$$\therefore E(X) = \mu = 3.5 = 21/6$$

Now

$$E(X^2) = \sum x^2 f(x)$$

$$= 1^2\times\frac{1}{6} + 2^2\times\frac{1}{6} + 3^2\times\frac{1}{6} + 4^2\times\frac{1}{6} + 5^2\times\frac{1}{6} + 6^2\times\frac{1}{6}$$

$$= 91/6$$

$$\therefore Var(X) = 91/6 - (21/6)^2$$

$$= 105/36$$

$$\therefore \boxed{Var(X) = 2.917}$$

Answer 2

(h) Given $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$

Now, given

$$E(X) = \int_0^\infty x f(x) = \int_0^\infty x \lambda e^{-\lambda x}\, dx$$

Now using integration by parts we get

$$= \lambda x \int e^{-\lambda x} dx - \int \lambda \frac{e^{-\lambda x}}{-\lambda} dx$$

$$= \left[-xe^{-\lambda x} - \frac{e^{-\lambda x}}{\lambda} \right]_0^\infty = 0 - \left(-\frac{1}{\lambda}\right)$$

$$\therefore \boxed{E(X) = \frac{1}{\lambda}}$$

Answer 2

(h) continued...

Now calculating $Var(X) = E(X^2) - (E[X])^2$

$$E(X^2) = \int_0^\infty x^2 f(x) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx$$

Now using Integration by parts

$$= \lambda x^2 \int \cancel{\lambda} e^{-\lambda x} dx - \int 2\lambda x \frac{e^{-\lambda x}}{-\lambda} dx$$

$$= \frac{\lambda x^2 e^{-\lambda x}}{-\lambda} + \frac{2}{\lambda} \int \lambda x e^{-\lambda x} dx$$

Again using Integration by parts we get

$$= -x^2 e^{-\lambda x} + 2\left[x \int e^{-\lambda x} dx - \int \frac{e^{-\lambda x}}{-\lambda} dx \right]$$

$$= -x^2 e^{-\lambda x} + 2\left[\frac{x e^{-\lambda x}}{-\lambda} - \frac{e^{-\lambda x}}{\lambda^2} \right]$$

$$= \left[-x^2 e^{-\lambda x} - \frac{2x e^{-\lambda x}}{\lambda} - \frac{2e^{-\lambda x}}{\lambda^2} \right]_0^\infty$$

Applying the limit we get

$$E(X^2) = \frac{2}{\lambda^2}$$

$$\therefore Var(X) = 2/\lambda^2 - (1/\lambda)^2$$

$$\therefore \boxed{Var(X) = \frac{1}{\lambda^2}}$$

(3) Continuous Distributions **(10 points)**

(a) Let us assume that the life of pen drives before failure is normally distributed with mean = 10 years and a standard deviation of 2 years. Find the probability that the pen drive fails between 9 years and 11 years. (**5 points**)

Answer

Given $\mu = 10$, $\sigma = 2$.

Using z transformation, we get

$z_1 = (11 - 10) / 2 = 0.5$

p(0.5) = 0.6915.

Similarly, for 9years we get,

$z_2 = (9 - 10) / 2 = -0.5$

p(-0.5) = 0.3085

Thus we have cumulative probability of failure up till 11 years and cumulative probability of failure up till 9 years. Thus probability of failure between 9 an 11 years is given by

Probability = 0.6915 – 0.3085 = **0.3830**

(b). Let us assume that your instructor assigns a letter grade of "Pass" to final score of 75 ± d. It is known that students' scores are normally distributed with mean of 75 and a standard deviation of 5. Find the value of d such that C's covers 95% of scores. (**5 points**)

Answer

Given $\mu = 75$, $\sigma = 5$, Confidence level = 95%.

PASS = 75 ± d.

Thus $\alpha = (1 - 95/100) = 0.05$

z-score (1-0.05/2) = z-score(0.0975) = 1.96.

now using z transform

$$z = \frac{x - \mu}{\sigma}$$

here $x - \mu = d$

So d/5 = 1.96

Thus d = 9.8 and **PASS = 75 ± 9.8**.

(4) Maximum Likelihood Estimation (MLE) (**10 points**)

(a) Concisely describe MLE procedure for single parameter **(2 points)**

Answer

Following is the general procedure for a single parameter MLE.

1. Write down the likelihood function L(Θ)
2. Maximize likelihood.
    a. Take log likelihood as product is difficult to maximize to get I(Θ)
    b. Differentiate I(Θ) w.r.t the parameter Θ
    c. Set derivative to 0 and solve the resulting equation.
    d. Check if this is maximum. Take the second derivative to check this. Generally not needed in case of uni-modal Gaussian distribution).

(b) Let X be a continuous random variable with p.d.f. for $\lambda > 0$, is defined as

$$f(x;\lambda) = \begin{cases} \lambda x^{\lambda-1} \mathit{if}\, 0<x<1 \\ 0\, \mathit{otherwise} \end{cases}$$

Then find MLE for the parameter λ. **(8 points)**

Answer

The solution to this question is present in the figures attached below.

Answer 4

(b) Given $f(x;\lambda) = \begin{cases} \lambda x^{\lambda-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

Step 1 : Define Likelihood function

$$L(\lambda) = \prod_{i=0}^{n} \lambda x_i^{\lambda-1}$$

Step 2 : Maximize Likelihood

Taking log on both sides will keep the estimate same since log is monotonic.

Hence maximizing log likelihood.

$$\therefore I(\lambda) = \log\left(\prod_{i=0}^{n} \lambda x_i^{\lambda-1}\right)$$

$$= \sum_{i=0}^{n} \log\left(\lambda x_i^{\lambda-1}\right)$$

$$= \sum_{i=0}^{n} \log\lambda + \sum_{i=0}^{n} (\lambda-1)\log x_i$$

$$= n\log\lambda + (\lambda-1)\sum_{i=0}^{n} \log x_i$$

$$= n\log\lambda + \lambda\sum_{i=0}^{n} \log x_i - \sum_{i=0}^{n} \log x_i$$

Answer 4

(b) continued . . .

Now differentiating w.r.t. λ we get

$$\frac{d\,I(\lambda)}{d\lambda} = \frac{n}{\lambda} + \sum_{i=0}^{n} \log x_i$$

Equating this to 0, we get

$$\frac{n}{\lambda} + \sum_{i=0}^{n} \log x_i = 0$$

$$\therefore \boxed{\lambda = -\frac{n}{\sum_{i=0}^{n} \log x_i}}$$

<u>(5) CLT, CI (**15 points**)</u>

(a) Define Central Limit Theorem and state assumptions (5 points)

Answer

Central Limit Theorem states that for a given population with finite mean, μ and finite variance, σ^2 the sampling distribution of mean approaches the normal distribution with mean same as population mean and variance as σ^2/N where N is the sample size.

The assumptions are as follows:

- Independence: Sample observations must be independent. In general, if sampling without replacement then, n < 10% of population.
- If case of skewed population or unknown population distribution the sample size should be large enough. In general n > 30.

(b) Define Confidence Interval for population mean (2 points)

Answer

Confidence Interval quantifies the quality of a particular estimate, which can also be population mean. The narrower the interval, the closer we are to the true value with higher probability.

(c) Find the critical value for 95% C.I. (1 point)

Answer

The critical value for 95% C.I is 1.96.

(d) Outline the procedure for finding C.I. (2 points)

Answer

The general outline for finding CI is as follows:

1. Identify the sample statistic (e.g. mean)
2. Select the confidence level (e.g. 90%, 95%,..)
3. Compute margin of error or
    a. critical value : The critical value can be calculated using CLT and can be expressed as z-score. Then this corresponding z-score can be converted to probability using z-table.
4. Finally C.I = Sample statistic ± margin of error.

(e) A sample of 30 students is drawn from CSC-522 population of 100. Average weight of sample is 150 pounds and standard deviation is 20 pounds. Compute the 95% CI. (5 points)

Answer

C.I can be calculated as follows:

Step1: Identify the sample statistic

μ_{sample} = 150, n = 30, s = 20;

Step2: Select Confidence level

Confidence level = 95%;

Step3: Compute margin of error.

α = (1 – 95/100) = 0.05

critical probability, p* = 1-0.05/2 = 0.975.

z-score(p*) = 1.96.

$$S.E = \frac{s}{\sqrt{n}} = \frac{20}{\sqrt{30}} = 3.65$$

margin of error = 1.96 * 3.65 = 7.154.

Thus C.I lies between **142.49** to **157.15**.

(6) Hypothesis testing fundamentals (**10 points**)

(a) Define null and alternate hypothesis (**1 points**)

Answer

Null hypothesis (H_0) is a statistical hypothesis used in inference that states that there is no difference between a specific value and the parameter or between two parameters. It normally refers to a general statement or the default position.

Alternate hypothesis (H_1) is the statistical hypothesis used in inference that states the existence a difference between a specific value and the parameter or between two parameters. This is normally the rival to the null hypothesis and is considered true when the null hypothesis is rejected.

(b) State null and alternate hypothesis for left--, right, and two-tailed tests (**1 points**)

Answer

The following table shows the values for Null and Alternate hypothesis for left, right and two tailed test. Here μ represents the mean which is the parameter under consideration and k represents the specific value.

Tests	Left	Right	Two - tailed
Null Hypothesis	$\mu = k$	$\mu = k$	$\mu = k$
Alternate Hypothesis	$\mu < k$	$\mu > k$	$\mu \neq k$

(c) State type-1 and type-2 errors (**1 points**)

Answer

Type-1 error occurs when the null hypothesis is rejected even when its actually true.

Type-2 error occurs when the null hypothesis is not rejected even when it is actually false.

The below table shows the complete summary of the type-1 and type-2 errors

	H_0 True	H_0 False
Reject H_0	Error (Type I)	Correct Decision
Do not Reject H_0	Correct Decision	Error (Type II)

(d) Define level of significance and list 3 widely accepted significance levels. (**1 points**)

Answer

Level of significance can be defined as the maximum probability of type 1 error. This probability is symbolized by α.

The 3 widely accepted significance levels are 0.10, 0.05 and 0.01.

(e) How do you reduce type-1 errors (**1 points**)

Answer

The number of type-1 errors are directly related to significance level. So if the significance level is 0.1 then there are 10% chances of having type-1 error, if the significance level is 0.05 then 5% chance of having a type-1 error. Thus as we lower the significance level the type-1 errors will also reduce.

(f) Define critical value (**1 point**)

Answer

Critical value can be defined as a value that separates critical region from non-critical region. Critical region in turn indicates the range of values of the test values that are significantly different from the actual value and thus for these values the null hypothesis should be rejected. Critical value separates these range of values from non-critical ones.

(g) Write down the general hypothesis testing procedure (**2 points**)

Answer

The general hypothesis testing procedure is as follows:

1. State the hypothesis and identify the claim.
2. Find the critical value(s) from the appropriate table.
3. Compute the test value.
4. Make the decision to reject or accept the null hypothesis.
5. Summarize the results.

(h) Write down z-test (**1 points**)

Answer

z-test is a statistical test for the mean of the population. It can be used when the sample size is greater than 30 and standard deviation of the population is given.

The formula for z-test is

$$z = \frac{(x - \mu)}{\sigma/\sqrt{n}}$$

where
x = sample mean
μ = hypothesized population mean
σ = population standard deviation
n = sample size.

(i) Which test do you use when population standard deviation is not known (**1 point**)

Answer

I will use t-test when population standard deviation is not known.

(7) Hypothesis testing (**10 points**)

Average weight of last year CS graduate students is 162.5lb, with a standard deviation of 6.9 lb. A sample of 50 students from this year batch is 165.2lb with same standard deviation. Answer the following: (1) Is there a reason to believe that there is a change in average weight of current batch of students? State your conclusion using traditional hypothesis testing using critical value of 0.05; (2) Using P-value method, state your conclusion.

Answer

(1) We follow the general hypothesis procedure for part (1). Since n >= 30 and standard deviation is given we opt for z-test.

Step1: Stating the hypothesis.

$H_0 : \mu = 162.5$

$H_1 : \mu \neq 162.5$

Step2: Defining critical value

Given $\alpha = 0.05$. Since this is a two-tailed test the critical value can be calculated as

C.V.(right) = z-value of (1- $\alpha/2$) = z-value (0.975) = 1.96

C.V.(left) = -1.96

Step3:

Using z-test formula we get

$$z = \frac{(165.2 - 162.5)}{6.9/\sqrt{50}} = 2.767$$

Step 4: Make decision

Since the z-test value is not within the C.V. Hence we reject the null hypothesis in the favor of alternate hypothesis.

Step 5: Summary

Hence there is enough evidence to support the claim that there is a change in the average weight of the current batch students compared to the previous batch.

(2) The cumulative probability of the computed z-value calculated from step 3 is

P(2.767) = 0.9972

Thus p-value can be calculated as 1 – 0.9972 = **0.0028.**

This p-value is much lesser than given α value which is 0.05. Thus p-value also provides the evidence that there is a change in the average weight of the current year students.

(8) R Mini-project (**20 points**) (Due: 10/11/15: 23.55pm)

Implement an R program to demonstrate CLT. At the minimum, your project should implement the following elements. Submit code (r-script as text file with reasonable documentation); if needed individuals will be asked to run their code (no modifications are allowed after submission).

(A) Should take **inputs**: (1) type of population distribution (e.g., uniform, normal, etc), (2) sample size (~30), (3) number of samples (>100).
(B) **Output**: Show the plot of sampling distribution