

Yusuf Keim
Ciger

85 %

Data Analysis

Technique
Assignment - 1

1)

1. Starting from the general expressions for the mean μ and variance σ^2 of a pdf $f(x)$, namely

$$\mu = \int xf(x) dx,$$

$$\sigma^2 = \int (x - \bar{x})^2 f(x) dx,$$

show that, for the uniform distribution $U(0, 1)$, these quantities are:

$$\mu = 0.5,$$

$$\sigma = \sqrt{\frac{1}{12}}.$$

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

[4]

→ To show for the uniform distribution with its mean and standard deviation, we should perform the integration with its lower limit $-\infty$, and upper limit $+\infty$.

$$\mu = \int_{-\infty}^{+\infty} x f(x) dx \Rightarrow \int_0^1 x dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2} \quad \checkmark \quad 1/1$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - 0.5)^2 f(x) dx \Rightarrow \int_0^1 (x - 0.5)^2 dx = \int_0^1 (x^2 - x + 0.25) dx$$

$$= \left[\frac{x^3}{3} - \frac{x^2}{2} + 0.25x \right]_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12} \Leftrightarrow \sigma = \sqrt{\frac{1}{12}} \quad \times \quad 2/2 \quad 1/1$$

2/2

2. A defence system is 99.5% efficient in intercepting ballistic missiles. (a) What is the probability that it will intercept all of 100 missiles launched against it? (b) How many missiles must an aggressor launch to have a better than evens (i.e. > 50%) chance of one or more penetrating the defences? (c) How many missiles would be needed to ensure a better than evens chance of more than two missiles evading the defences? (Hint: Do you think the number would be large enough that you could consider an approximation?)

[10]

a) Let X be the number of missiles intercepted, that follows a binomial distribution

$$X \sim \text{Binomial}(100, 0.995)$$

$$P(X=100) = p^{100} = (0.995)^{100}$$

$$= \approx 0.6058 = 60.58\%$$

✓ ✓ ✓

b) $P(\text{at least one penetrates}) = 1 - p^n$

And we need,

$$1 - p^n > 0.5$$

✓ ✓ ✓

Sigma
flips since $\therefore p^n < 0.5$

log is $\therefore n \ln(p) < \ln(0.5)$

$n \ln(0.995) < \ln(0.5)$

$n < \frac{\ln(0.5)}{\ln(0.995)} \Rightarrow n > \frac{-0.69}{-0.005} \approx 138.27$

✓ ✓ ✓

c) $n = \underline{139}$

✓ ✓ ✓

let Y be the number of missiles that evade interception.

$Y \rightarrow \text{Binomial Distribution}$. We want $P(Y \geq 2) > 0.5 \Rightarrow 1 - P(Y \leq 2) > 0.5$

$$\Rightarrow P(Y \leq 2) < 0.5 \Rightarrow [P(X=0) + P(X=1) + P(X=2)] < 0.5$$

✓ ✓ ✓

$$P(X=0) = \binom{n}{0} p^0 (1-p)^n = (0.995)^n$$

$$P(X=1) = \binom{n}{1} p^1 (1-p)^{n-1} = (0.005) \cdot (0.995)^{n-1}$$

$$P(X=2) = \binom{n}{2} p^2 (1-p)^{n-2} = \frac{n(n-1)}{2} \cdot (0.005)^2 \cdot (0.995)^{n-2}$$

$$\Rightarrow [(0.995)^n + [0.005] \cdot (0.995)^{n-1}] + [\frac{n(n-1)}{2} \cdot (0.005)^2 \cdot (0.995)^{n-2}] < 0.5$$

Could also use poisson stat.

2 / 3

To find the value of n , I am going to provide a Python script

```
1 from scipy.stats import binom
2
3 p = 0.005
4 threshold_prob = 0.5
5 n = 1
6
7 while True:
8     prob_X_leq_2 = binom.cdf(2, n, p)
9
10    if prob_X_leq_2 < threshold_prob:
11        break
12    n += 1
13
14 print("The smallest number of missiles needed is:", n)
✓ 0.5s
The smallest number of missiles needed is: 535
```

As we can see from the output we need at least $n = 535$ missiles.

✓ 2 / 2

3) For this question, below, I add the screenshot of my Python notebook

A) Finding the statistics of data

```
1 data = pd.read_csv('ccd_image.txt', header = None, delimiter=',')
2
3 data.head()
0 1279 1244 1285 1283 1329 1303 1296 1297 1292 1232 ... 1146 1190 1116 1180 1208 1137 1169 1132 1141 1137
1 1301 1334 1220 1313 1329 1285 1295 1262 1258 1308 ... 1207 1123 1158 1150 1193 1168 1131 1229 1126 1156
2 1352 1350 1299 1255 1276 1267 1265 1366 1297 1327 ... 1177 1195 1192 1145 1194 1142 1181 1175 1147 1127
3 1267 1303 1323 1368 1243 1298 1332 1277 1316 1314 ... 1192 1214 1139 1161 1143 1154 1136 1153 1128 1153
4 1267 1274 1324 1293 1294 1328 1258 1258 1320 1303 ... 1171 1206 1192 1134 1194 1195 1169 1181 1166 1195
5 rows × 255 columns
```

B) Plotting the data

```
1 plt.figure(figsize=(10, 6))
2 plt.hist(values, bins=100, range=(1000, 3000), log=True, color='skyblue', edgecolor='black')
3 plt.xlabel('ADU')
4 plt.ylabel('Frequency (log scale)')
5 plt.title('Histogram of CCD Image Data (1000-3000 ADU)')
6 plt.show()
```

✓ 2 / 2

Histogram of CCD Image Data (1000-3000 ADU)

Frequency (log scale)

ADU

C)

- If we just checked the histogramic data, we observed that 'Median' is actually our best candidate.
- It is because of sensitivity to extreme values. We can also observe that 'Median' provides the central tendency of data that is not influenced by the star brightness levels

justify more.

MEAN is biased by large outliers.

MODE is noisy and > MEAN

✓ 1 / 2

An experiment carried out twice produces the following results:

$$12 \pm 2$$
$$9 \pm 3.$$

Calculate the weighted mean result and its uncertainty. What are the 90% confidence limits on the weighted mean? (By x% confidence limits we mean the (usually symmetric) boundaries within which x% of the area falls. Hint: $\frac{1}{\sqrt{2\pi}} \int_{-1.645}^{1.645} \exp(-z^2/2) dz = 0.90$.)

We know that $\mu_1 = 12, \sigma_1 = 2$ and $\mu_2 = 9, \sigma_2 = 3$

↳ To find the weighted mean, I am going to use this formula: $\bar{x}_{\text{weighted}} = \frac{\sum \left(\frac{x_i}{\sigma_i^2} \right)}{\sum \left(\frac{1}{\sigma_i^2} \right)}$

$$\bar{x}_{\text{weighted}} = \frac{\left(\frac{12}{4} + \frac{9}{9} \right)}{\left(\frac{1}{4} + \frac{1}{9} \right)} = \frac{4}{\frac{13}{36}} = 4 \cdot \frac{36}{13} \approx 11.08$$

✓

↳ To find the uncertainty in weighted mean: $\sigma_{\bar{x}_{\text{weighted}}} = \sqrt{\sum \left(\frac{1}{\sigma_i^2} \right)}$

$$= \sqrt{\frac{1}{\frac{13}{36}}} \approx 1.66$$

✓

2 / 2

↳ To determine the 90% Confidence Interval by: $\bar{x}_{\text{weighted}} \pm 1.645 \times \sigma_{\bar{x}_{\text{weighted}}}$.

$$\Rightarrow (11.08 - 1.645 \times 1.66, 11.08 + 1.645 \times 1.66)$$

$$\Rightarrow (8.34, 13.81)$$

✓

1 / 1

5)

When a muon decays, an electron is emitted at an angle θ relative to the spin, such that the probability density function of the angular distribution is

$$dP = k(1 + \alpha x) dx,$$

where $x = \cos \theta$ and α is a constant that must be determined experimentally. Find the value of the constant k , and the mean and variance of the x distribution (in terms of α). [5]

↳ To normalize the PDF, we need the total probability over the range of $\cos \theta(x)$ which is -1 to 1 to be equal to 1 .

$$\Rightarrow \int_{-1}^1 k dx + \int_{-1}^1 k \alpha x dx \Rightarrow k[x]_{-1}^1 + k \alpha [\frac{x^2}{2}]_{-1}^1 = 1$$

$$\Rightarrow 2k + k\alpha 0 = 1 \Leftrightarrow 2k = 1 \quad k = \frac{1}{2}, \quad \checkmark 2/2$$

$$\text{so } dP = \frac{1}{2}(1 + \alpha x) dx$$

↳ To find the mean:

$$\mu = \int_{-1}^1 x \cdot \frac{1}{2}(1 + \alpha x) dx = \int_{-1}^1 \underbrace{\frac{1}{2}x dx}_{0} + \underbrace{\frac{\alpha}{2} \int_{-1}^1 x^2 dx}_{1}$$

$$= \frac{\alpha}{2} \left[\frac{x^3}{3} \right]_{-1}^1 = \frac{\alpha}{2} \cdot \frac{2}{3} = \boxed{\frac{\alpha}{3}} = \mu \quad \checkmark 1/1$$

↳ To calculate the $\text{var}(x) = \langle x^2 \rangle - \langle x \rangle^2$

$$\langle x^2 \rangle = \int_{-1}^1 x^2 \frac{1}{2}(1 + \alpha x) dx = \frac{1}{2} \int_{-1}^1 x^2 dx + \frac{\alpha}{2} \int_{-1}^1 x^3 dx$$

$$= \frac{1}{2} \left[\frac{x^3}{3} \right]_{-1}^1 + \frac{\alpha}{2} \left[\frac{x^4}{4} \right]_{-1}^1 = \frac{1}{3} + \frac{\alpha}{2} \cdot 0 = \frac{1}{3}$$

$$\langle x \rangle^2 = \frac{\alpha^2}{9}, \text{ so } \text{var}(x) = \frac{1}{3} - \frac{\alpha^2}{9} = \boxed{\frac{3 - \alpha^2}{9} = \text{Var}(x)}$$

6)

6. This question is a simple exercise in changing variables. A gun fires at random in the angular range $-\pi/2 < \theta < \pi/2$ towards a wall a distance l away. If y is the coordinate along the wall, show that

$$g(y)dy = \frac{1}{\pi} \frac{1}{1 + (y/l)^2} \frac{dy}{l}.$$

This, as we saw in the lectures, is the Cauchy distribution. Assuming that the mean should be zero from symmetry considerations, try to find the standard deviation; what problem do you have? Truncate the distribution at a distance $|y| = L$ (not the same as l) either side of the peak. Calculate the new normalisation constant, and find the standard deviation. [12]

so, $y = l \tan \theta$, so the range of y is $(-\frac{\pi}{2}, \frac{\pi}{2})$

$$\frac{dy}{d\theta} = l \sec^2 \theta \Rightarrow dy = l \sec^2 \theta d\theta, d\theta = \frac{dy}{l \sec^2 \theta}$$
✓ 1/1

- Expressing $g(y)dy$ in terms of θ : we already know that pdf $f(\theta)$ for θ has intervals of $(-\frac{\pi}{2}, \frac{\pi}{2})$. means that $f(\theta)d\theta = \frac{1}{\pi} d\theta$

$$f(\theta)d\theta = \frac{1}{\pi} \cdot \frac{1}{l \sec^2 \theta} dy$$

we know that $\sec^2 \theta \Rightarrow 1 + \tan^2 \theta$, and $\tan \theta = \frac{y}{l}$

$$\hookrightarrow f(\theta)d\theta = \frac{1}{\pi} \cdot \frac{1}{l} \cdot \frac{l}{1 + (\frac{y}{l})^2} dy$$

$$\hookrightarrow g(y)dy = \frac{1}{\pi} \cdot \frac{1}{l} \cdot \frac{1}{1 + (\frac{y}{l})^2} dy$$
✓ 2/2

- Finding standard deviation assuming $\mu=0$.

$$\sigma^2 = \int_{-\infty}^{+\infty} (y - \mu)^2 g(y) dy \Rightarrow \int_{-\infty}^{\infty} y^2 \cdot \frac{1}{\pi l} \cdot \frac{1}{1 + (\frac{y}{l})^2} dy$$

$$\Rightarrow \frac{1}{\pi l} \cdot \int_{-\infty}^{+\infty} \frac{y^2}{1 + (\frac{y}{l})^2} dy$$

→ After this point, since the Cauchy Distribution are too heavy? the integral for the variance diverges.

✓ 3/3

↳ Truncate distribution

↳ $\int_{-L}^L g(y) dy = 1$, substitute $g(y) = \frac{c}{\pi \cdot e} \cdot \frac{1}{1 + (\frac{y}{e})^2}$, $c \Rightarrow \text{normalization constant.}$

↳ $\int_{-L}^L \frac{c}{\pi e} \cdot \frac{1}{1 + (\frac{y}{e})^2} dy = 1 \Leftrightarrow \frac{c}{\pi e} \int_{-L}^L \frac{1}{1 + (\frac{y}{e})^2} dy = 1$

$$c = \pi e \int_{-L}^L \left(\frac{1}{1 + (\frac{y}{e})^2} \right)^{-1} dy$$



$$1 = \frac{c}{\pi e} \left[\arctan \left(\frac{y}{e} \right) \right]_{-L}^L \Rightarrow$$

$$\Rightarrow c = \frac{\pi}{2 \arctan(\frac{L}{e})}$$

0/3

Missing σ^2 vs L calculation

0/3

7. A simple example of changing variables.

7)

(a) Let x_1, x_2 be independent random variables. Let

$$\left. \begin{array}{l} u_1 = x_1 - x_2 \\ u_2 = x_1 + x_2. \end{array} \right\} \quad u = A \cdot x \Rightarrow u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} := A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

Assume we are given

$$V = \begin{pmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{pmatrix}.$$

Find V' , the error matrix for u_1 and u_2 . Substitute in $\sigma_{x_1} = 5$, $\sigma_{x_2} = 3$, and evaluate V' , and hence σ_{u_1} and σ_{u_2} .

$$A^T = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

1/1

(b) Undoing the error propagation:

$$\begin{aligned} x_1 &= \frac{u_1 + u_2}{2}, \\ x_2 &= \frac{u_2 - u_1}{2}. \end{aligned}$$

What do you get for $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ (using the numerical values for σ_{u_1} and σ_{u_2} found above) if you ignore correlations?

Now use the correct matrix method to recover the expected

$$V = \begin{pmatrix} 25 & 0 \\ 0 & 9 \end{pmatrix}.$$

1/1

a) $V = \begin{pmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{pmatrix} \quad V' = A \cdot V \cdot A^T$

$$A \cdot V = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{x_1}^2 & -\sigma_{x_2}^2 \\ \sigma_{x_1}^2 & \sigma_{x_2}^2 \end{pmatrix}$$

$$A \cdot V \cdot A^T = \begin{pmatrix} \sigma_{x_1}^2 & -\sigma_{x_2}^2 \\ \sigma_{x_1}^2 & \sigma_{x_2}^2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} \sigma_{x_1}^2 + \sigma_{x_2}^2 & \sigma_{x_1}^2 - \sigma_{x_2}^2 \\ \sigma_{x_1}^2 - \sigma_{x_2}^2 & \sigma_{x_1}^2 + \sigma_{x_2}^2 \end{pmatrix}$$

2/2

Now just replace $\sigma_{x_1} = 5$ and $\sigma_{x_2} = 3$

$$V' = \begin{pmatrix} 34 & 16 \\ 16 & 34 \end{pmatrix} \quad \text{The variances for } u_1 \text{ and } u_2$$

Now $\text{Var}(u_1) = \text{Var}(u_2) = 34$

1/1

$\sigma_{u_1} = \sigma_{u_2} = \sqrt{34} \times$

b) Ignoring Correlations.

We already know that $\sigma_{u_1}^2 = \sigma_{u_2}^2 = 34$, and by ignoring correlations we assume $\text{Cov}(u_1, u_2) = 0$.

$$\sigma_{x_1}^2 = \left(\frac{1}{2}\right)^2 \sigma_{u_1}^2 + \left(\frac{1}{2}\right)^2 \sigma_{u_2}^2 \Rightarrow \frac{1}{4} \cdot 34 + \frac{1}{4} \cdot 34 = 17$$

2/2

$\sigma_{x_2}^2$ = By using same formula , $\text{var}(x_2) = \text{var}(x_1) = 17.$

$$\begin{pmatrix} 17 & 0 \\ 0 & 17 \end{pmatrix} \neq \begin{pmatrix} 25 & 0 \\ 0 & 9 \end{pmatrix}$$

- Correct Matrix Method.

$$V' = \begin{pmatrix} \sigma_{u_1}^2 & \text{Cov}(u_1, u_2) \\ \text{Cov}(u_1, u_2) & \sigma_{u_2}^2 \end{pmatrix} = \begin{pmatrix} 34 & 16 \\ 16 & 34 \end{pmatrix}$$

$$, \kappa_2 = \frac{u_2 - u_1}{2} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

$$V = A^{-1} V' (A^{-1})^T$$

$$A^{-1} V' = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 34 & 16 \\ 16 & 34 \end{pmatrix} = \begin{pmatrix} 25 & 25 \\ -9 & 9 \end{pmatrix}$$

$$V = \begin{pmatrix} 25 & 25 \\ -9 & 9 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 25 & 0 \\ 0 & 9 \end{pmatrix} \Rightarrow \text{By correct matrix method, we find the equivalent.}$$

✓ 1/1