Experimental Statistics for Engineers I

Kamala Dadashova

Problem 1: Let X have a pdf given by

$$f(x) = \begin{cases} \frac{\theta}{x^2} & x \geq \theta \\ 0 & otherwise \end{cases}$$

(a) Show that this is a valid pdf.

Solution:

Firstly, $f(x) \geq 0$ holds true trivially. Secondly, we have to show $\int_{-\infty}^{\infty} f(x)\ dx = 1$.

$$\int_{-\infty}^{\infty} f(x)\ dx = \int_{\theta}^{\infty} \frac{\theta}{x^2}\ dx = \lim_{t \to \infty} \int_{\theta}^{t} \frac{\theta}{x^2}\ dt = \lim_{t \to \infty} -\frac{\theta}{x}|_{\theta}^{t} = 1$$

Since both conditions hold, it is a valid pdf.

(b) Find the CDF of X.

Solution:

If $x < \theta$, then cdf, F(x)=0. If $x \geq \theta$, then cdf

$$F(x) = \int_{\theta}^{x} \frac{\theta}{t^2}\ dt = \frac{-\theta}{t}|_{\theta}^{x} = \frac{-\theta}{x} + 1.$$

$$F(x) = \begin{cases} 0 & x < \theta \\ -\frac{\theta}{x} + 1 & x \geq \theta \end{cases}$$

(c) Find the 30^{th} percentile of the distribution.

Solution:

$$p = F(\eta(p))$$

$$0.3 = -\frac{\theta}{\eta(p)} + 1$$

$$\frac{\theta}{\eta(p)} = 0.7$$

$$\eta(0.3) = \frac{\theta}{0.7}$$

(d) Find the general form for $P(X > \theta + b)$ where b is a positive constant.

Solution:

$$P(X > \theta + b) = 1 - F(\theta + b) = 1 - (-\frac{\theta}{\theta + b} + 1) = \frac{\theta}{\theta + b}.$$

(e) If c and d are both positive constants such that $d > c$, find $P(X > \theta + d|X > \theta + c)$

Solution:

We use memory-less property which is

$$P(X \geq t + t_0 | X \geq t_0) = P(X \geq t)$$

In this case, our $t_0 = \theta + c$ and $t = d - c$. Using above property, we get

$$P(X > d - c) = 1 - F(d - c) = 1 - (-\frac{\theta}{d - c} + 1) = \frac{\theta}{d - c}$$

Here, we do not know whether $d - c > \theta$ or no. So, we will have another case,

$$P(X > d - c) = 1 - F(d - c) = 1 - (0) = 1.$$

(f) Find the expected value of $X^{\frac{1}{2}}$.

$$E(X^{\frac{1}{2}}) = \int_{\theta}^{\infty} x^{\frac{1}{2}} \frac{\theta}{x^2}\, dx = \lim_{t \to \infty} \int_{\theta}^{t} \frac{\theta}{x^{1.5}}\, dt = -2 \lim_{t \to \infty} \frac{\theta}{x^{0.5}}|_{\theta}^{t} = 2\sqrt{\theta}$$

Problem 2: A software development company has three jobs to do. Two of the jobs require three programmers, and the other requires four. If the company employs ten programmers, how many ways are there to assign them to the jobs?

Solution:

We are given that company employs 10 programmers. Since two of the jobs require three programmers and the third job require four jobs, we have 3 programmers left out of 10. Thus, the number of ways to assign them is equal to

$$\binom{10}{4} \times \binom{6}{3} \times \binom{3}{3} = 4200$$

Problem 3: Suppose that n components are connected in series. For each unit, there is backup unit, and the system fails if and only if both a unit and its backup fail. Assuming that all the units are independent and fail with probability p, what is the probability that the system works (find and expression in terms of n and p)?

Solution:

The system fails iff both units fail. Probability for the failure is p. Since the units are independent, $p \times p = p^2$. Hence, the probability that a component works is $(1 - p^2)$. All n components have to work implying that the probability for the system to works is $(1 - p^2)^n$.

Problem 4: A factory runs three shifts. In a given day, 1% of the items produced by the first shift are defective, 2% of the second shift's items are defective, and 5% of the third shift's items are defective. If an item is defective, what is the probability that it was produced by the third shift? Assume an item is equally likely to come from each shift.
Solution:
Let denote F for the first shift, S- for the second shift, T- for the third shift, D is for the defective items. We are given the following facts; $P(F) = P(S) = P(T) = \frac{1}{3}$ and $P(D|F) = 0.01$, $P(D|S) = 0.02$, $P(D|T) = 0.05$. We are required to find $P(T|D)$. Using Bayes' Theorem, we obtain

$$P(T|D) = \frac{P(D|T)P(T)}{P(D|F) \times P(F) + P(D|S) \times P(S) + P(D|T) \times P(T)} = \frac{0.05 \times \frac{1}{3}}{\frac{1}{3}(0.01 + 0.02 + 0.05))} = \frac{5}{8}$$

Problem 5: Let X and Y have joint pdf

$$f_{X,Y}(x, y) = \begin{cases} 1 & 0 \leq x \leq 2, 0 \leq y \leq 1, 2y \leq x \\ 0 & otherwise \end{cases}$$

(a) Show that this is a valid pdf.

Solution:

First condition $f(x, y) \geq 0$ holds. The second condition is $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\ dy\ dx = 1$. Let us show it,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\ dy\ dx = \int_{0}^{2} \int_{0}^{0.5x} 1\ dy\ dx = \frac{x^2}{4}|_0^2 = 1.$$

So, it is a valid pdf.

(b) Find $P(0.5 < x < 2, 0.5 < y < 1))$

Solution:

$$P(0.5 < x < 2, 0.5 < y < 1)) = \int^{2} \int_{0.5}^{0.5x} 1\ dy\ dx = \int_{0.5}^{2} 0.5x - 0.5\ dx = (0.25x^2 - 0.5x)|_{0.5}^2 = \frac{3}{16}$$

(c) Find the marginal distribution of Y.

Solution:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\ dx = \int_{2y}^{2} 1\ dx = 2 - 2y$$

$$f_Y(y) = \begin{cases} 2 - 2y & 0 \leq y \leq 1 \\ 0 & otherwise \end{cases}$$

(d) Find the conditional distribution of $X|Y$. What well-known distribution is this?

Solution:

$$f_{X|Y}(X|Y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{2-2y}$$

Since it is constant with respect to x, we say $X|Y \sim Unif(0, 2-2Y)$

(e) Write an expression for $Cov(X,Y)$ (You do not need to evaluate any integrals, just set up the expression to be evaluated).

Solution:

$$Cov(X,Y) = \mathbf{E}[(X-\mu_x)(Y-\mu_y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f(x,y)\ dx\ dy$$

Using part (c) and the formula for mean, we get

$$\mu_y = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \int_0^1 y(2-2y) = y^2 - \frac{2y^3}{3}|_0^1 = \frac{1}{3}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\ dy = \int_0^{0.5x} 1\ dy = \frac{x}{2}$$

$$f_Y(y) = \begin{cases} \frac{x}{2} & 0 \le x \le 2 \\ 0 & otherwise \end{cases}$$

$$\mu_x = \int_{-\infty}^{\infty} x f_X(x)\ dx = \int_0^2 x0.5x\ dx = \frac{x^3}{6}|_0^2 = \frac{4}{3}$$

$$Cov(X,Y) = \int_0^2\int_0^{0.5x}(x-\frac{4}{3})(y-\frac{1}{3})\ dy\ dx$$

Problem 6: A large lot of manufactured items contains 10% with exactly one defect, 5%with more than one defect, and the remainder with no defects. Ten items are randomly selected from this lot for sale. Find the probability of obtaining a random sample having 3 items with exactly one defect, 0 items with more than one defect, and 7 with no defects.
Solution:
Here, $n = 10,\ x_1 = 3,\ x_2 = 0,\ x_3 = 7.$

$$P(x_1,x_2,x_3) = \frac{10!}{3!0!7!}(0.1)^3 \times (0.05)^0 \times (1-0.15)^7 \approx 0.0385$$

Problem 7: A fair coin is tossed three times. Let X be the number of heads on the first toss and Y be the total number of heads.

(a) Create a table that describes the joint pmf.

Solution:

Let S be sample represents all possible outcomes,

$$\{TTT, HTT, THT, TTH, \ , HHT, \ , HTH, \ , THH, HHH\}$$

On the first toss, two cases might happen either X=0 (no head) or X=1 (head).

p(x,y)	Y=0	Y=1	Y=2	Y=3	
X=0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{4}{8}$
X=1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{4}{8}$
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

(b) Find the conditional distribution of Y given X= 0.

Solution:

$$P(Y = 0|X = 0) = \frac{p(0,0)}{p_X(0)} = \frac{\frac{1}{8}}{\frac{4}{8}} = \frac{1}{4}$$

$$P(Y = 1|X = 0) = \frac{p(0,1)}{p_X(0)} = \frac{\frac{2}{8}}{\frac{4}{8}} = \frac{1}{2}$$

$$P(Y = 2|X = 0) = \frac{p(0,2)}{p_X(0)} = \frac{\frac{1}{8}}{\frac{4}{8}} = \frac{1}{4}$$

$$P(Y = 3|X = 0) = \frac{p(0,3)}{p_X(0)} = \frac{0}{\frac{4}{8}} = 0$$

(c) Are X and Y independent? Support your answer.

Solution:

We will give a counterexample.

$$p(0,0) = \frac{1}{8} \neq p_X(0)p_Y(0) = \frac{4}{8} \times \frac{1}{8} = \frac{1}{16}$$

So, they are not independent.

Problem 8: Appending three extra bits to a 4-bit word in a particular way (a Hamming code) allows detection and correction of up to one error in any of the bits. Consider an example where each bit has probability 0.05 of being changed during communication, and the bits are changed independently of each other.

(a) What is the probability that the word is correctly received (that is, 0 or 1 bit is in error)?

Solution:

Let X be number of bits in error. We want to find $P(X=0)+P(X=1)$ and we use binomial distribution because of the structure of the problem ,

$$dbinom(x=0, size=7, prob=0.05) + dbinom(x=1, size=7, prob=0.05) = 0.9556$$

(b) How does this probability compare to the probability that the word will be transmitted correctly with no check bits, in which case all four bits would have to be transmitted correctly for the word to be correct?

Solution:

Now, we have n=4. We want to find $P(X=0)$.

$$dbinom(x=0, size=4, prob=0.05) = 0.8145$$

Problem 9: If two loads are applied to a cantilever beam as shown in the accompanying drawing, the bending moment at 0 due to the loads is $a_1X_1 + a_2X_2$

(a) Suppose that X_1 and X_2 are independent rvs with means 2 and 4 kips, respectively, and standard deviations 0.5 and 1.0 kip, respectively. If $a_1 = 5ft$ and $a_2 = 10ft$, what is the expected bending moment and what is the standard deviation of the bending moment?

Solution:

We want to find

$$\mathbf{E}(5X_1 + 10X_2) = 5\mathbf{E}(X_1) + 10\mathbf{E}(X_2) = 5 \times 2 + 10 \times 4 = 50$$

$$V(5X_1 + 10X_2) = 25V(X_1) + 100V(X_2) = 25 \times 0.25 + 100 \times 1 = 106.25$$
$$SD(5X_1 + 10X_2) = \sqrt{V(5X_1 + 10X_2)} \approx 10.3078$$

(b) If X_1 and X_2 are normally distributed, what is the probability that the bending moment will exceed 75 kip-ft?

Solution:

We want to find

$$P(5X_1 + 10X_2 > 75) = pnorm(q = 75, mean = 50, sd = 10.3078, lower.tail = F) = 0.0076$$

(c) If the situation is as described in part (a) except that $Corr(X_1, X_2) = 0.5$(so that the two loads are not independent), what is the variance of the bending moment?

Solution:

We will use formula for not independent case,

$$V(5X_1 + 10X_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_i a_j Cov(X_i, X_j)$$

Let us find $Cov(X_1, X_2) = \rho(X_1, X_2) \times \sigma_{X_1} \times \sigma_{X_2} = 0.5 \times 0.5 \times 1 = 0.25$, $Cov(X_1, X_1) = \rho(X_1, X_1) \times \sigma_{X_1} \times \sigma_{X_1} = 1 \times 0.5^2 = 0.25$ and $Cov(X_2, X_2) = \rho(X_2, X_2) \times \sigma_{X_2} \times \sigma_{X_2} = 1 \times 1^2 = 1$

Thus,

$$V(5X_1 + 10X_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_i a_j Cov(X_i, X_j) =$$

$$a_1^2 Cov(X_1, X_1) + a_1 a_2 Cov(X_1, X_2) + a_2 a_1 Cov(X_2, X_1) + a_2^2 Cov(X_2, X_2) =$$

$$25 \times 0.25 + 50 \times 0.25 + 50 \times 0.25 + 100 \times 1 = 131.25$$

Problem 10: Given the discrete uniform population

$$f(x) = \begin{cases} \frac{1}{3} & x = 2, 4, 6 \\ 0 & otherwise \end{cases}$$

find the probability that a random sample of size 54, selected with replacement, will yield a sample mean greater than 4.1 but less than 4.4.

Solution:

Here, we use CLT. Note that n=54. We firstly want to find sample mean and variance. Using CLT, $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$. For discrete rv,

$$\mu = \sum_x xp(x) = \frac{1}{3} \times 2 + \frac{1}{3} \times 4 + \frac{1}{3} \times 6 = 4$$

$$\sigma^2 = E(x^2) - \mu^2 = (4 \times \frac{1}{3} + 16 \times \frac{1}{3} + 36 \times \frac{1}{3}) - 4^2 = \frac{8}{3}$$

Thus,

$$\mu_{\bar{X}} = \mu = 4$$

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \frac{4}{81}$$

So, we want to find

$$P(4.1 \leq \bar{X} \leq 4.4) = pnorm(q = 4.4, mean = 4, sd = 2/9) - pnorm(q = 4.1, mean = 4, sd = 2/9) = 0.2904$$

Problem 11: According to a survey by the Administrative Management Society, one-half of U.S. companies give employees 4 weeks of vacation after they have been with the company for 15 years. Find the probability that among 6 companies surveyed at random, the number that give employees 4 weeks of vacation after 15 years of employment is

(a) Between 2 and 5 inclusive.

Solution: Let X be the random variable which is the number of companies that give employees 4 weeks of vacation after 15 years of employment. We will use binomial distribution for this problem. Note that n=6, p=0.5. We are required to find

$$P(2 \leq X \leq 5) = pbinom(q = 5, size = 6, prob = 0.5) - pbinom(q = 1, size = 6, prob = 0.5) = 0.875$$

(b) Fewer than 3.

Solution: Here, we want to find $P(X < 3)$. To do it, we first calculate $P(X \leq 3)$ then subtract $P(X = 3)$ from it.

$$P(X \leq 3) = pbinom(q = 3, size = 6, prob = 0.5) = 0.65625$$
$$P(X = 3) = dbinom(x = 3, size = 6, prob = 0.5) = 0.3125$$
$$P(X < 3) = 0.65625 - = 0.3125 = 0.34375$$

Problem 12: The life, in years, of a certain type of electrical switch has an exponential distribution with an average lifetime of 2 years.

(a) If 100 of these switches are installed in different systems, what is the probability that at most 30 fail during the first year?

Solution:

Let X be rv measures the time that a switch will happen. We have exponential distribution, $\mu = \frac{1}{\lambda}$. Thus, $\lambda = \frac{1}{2}$. The probability of failure during the first year,

$$P(X \leq 1) = pexp(q = 1, rate = 1/2) = 0.3935$$

Next, we have 100 of these switches installed in different systems. Let Y be the random variable measuring the the probability, which exactly k switches will fail in the first year.

$$P(Y \leq 30) = pbinom(q = 30, size = 100, prob = 0.3935) = 0.03343$$

(b) For a single system, what is the probability we can keep the switching mechanism operable for 10 years if we assume we will have to replace a switch due to failures three times during the 10-year period.

Solution:

We will have three failures during 10 years and it requires t_1, t_2, t_3, t_4 times. Our goal is to find $P(t_1 + t_2 + t_3 + t_4 > 10)$. Let $Y = t_1 + t_2 + t_3 + t_4$. We have $t_i \sim exp(\frac{1}{2})$ and using class note it is equivalent to $Y \sim Gamma(4, 2)$. Thus,

$$P(Y > 10) = pgamma(q = 10, shape = 4, scale = 2, lower.tail = F) = 0.2650$$

Problem 13: A manufacturing company uses an acceptance scheme on items from a production line before they are shipped. The plan is a two-stage one. Boxes of 25 items are readied for shipment, and a sample of 3 items is tested for defectives. If any defectives are found, the entire box is sent back for 100% screening. If no defectives are found, the box is shipped

(a) What is the probability that a box containing 3 defectives will be shipped?

Solution:

The box will be shipped if a sample of 3 items has no defective. Thus,

$$P = \frac{\binom{3}{0}\binom{22}{3}}{\binom{25}{3}} \approx 0.6696$$

Alternative solution would be using hyper geometric distribution

$$dhyper(x = 0, m = 3, n = 25 - 3, k = 3) \approx 0.6696$$

(b) What is the probability that a box containing only 1 defective will be sent back for screening?

Solution:

This box will be sent back for screening if a sample of 3 items has 1 defective and 2 non-defective items.

$$P = \frac{\binom{3}{1}\binom{22}{0}}{\binom{25}{1}} = 0.12$$

Alternatively,

$$dhyper(x = 1, m = 1, n = 25 - 1, k = 3) = 0.12$$

Problem 14: For a certain type of copper wire, it is known that, on the average, 1.5 flaws occur per millimeter. Assume that the number of flaws is a Poisson random variable.

(a) What is the probability that no flaws occur in a certain portion of wire of length 5 millimeters?

Solution: Let X be number of flaws. We use Poisson distribution with $\lambda = 1.5$ Since length is 5 millimeters, $\lambda = 1.5 \times 5 = 7.5$. We need to find no flaws case, i.e, P(X=0) with rate.

$$dpois(x = 0, lambda = 7.5) = 0.00055$$

(b) What is the mean number of flaws in a portion of length 5 millimeters?

Solution:

Using class note, we know expected or mean value for Poisson distribution is equal to μ. In this problem $\mu = 1.5 \times 5 = 7.5$.

Problem 15: The tensile strength of a certain metal component is normally distributed with a mean of 10,000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter.

(a) What proportion of these components exceed 10,150 kilograms per square centimeter in tensile strength?

Solution: Here, we have normal distribution with mean and standard deviation. We are required to find

$$P(X > 10150) = pnorm(10150, 10000, 100, lower.tail = F) = 0.0668$$

(b) If specifications require that all components have tensile strength between 9800 and 10,200 kilograms per square centimeter inclusive, what proportion of pieces would we expect to scrap?

Solution:

$$P(9800 \leq X \leq 10200) = pnorm(10200, 10000, 100) - pnorm(9800, 10000, 100) = 0.9545$$

Problem 16: If the proportion of a brand of television set requiring service during the first year of operation is a random variable having a beta distribution with $\alpha = 3$ and $\beta = 2$, what is the probability that at least 80% of the new models of this brand sold this year will require service during their first year of operation?

Solution: Let X be he proportion of a brand of television set requiring service during the first year of operation . We use beta distribution considering the given two parameters, $\alpha = 3$ and $\beta = 2$.

$$pbeta(q = .8, shape1 = 3, shape2 = 2, lower.tail = F) = 0.1808$$

CODES

Problem 8 (R code part)

(a)

```
dbinom(x=0,size=7,prob=0.05)+dbinom(x=1,size=7,prob=0.05)
```

```
## [1] 0.9556195
```

(b)

```
dbinom(x=0,size=4,prob=0.05)
```

```
## [1] 0.8145062
```

Problem 9:(R code part (c))

```
pnorm(q=75,mean=50,sd=10.3078,lower.tail=F)
```

```
## [1] 0.007646864
```

Problem 10:(R code part)

```
pnorm(q=4.4,mean=4,sd=2/9) -pnorm(q=4.1,mean=4,sd=2/9)
```

```
## [1] 0.2904249
```

Problem 11: (R code part) (a)

```
pbinom(q=5,size=6,prob=0.5) -pbinom(q=1,size=6,prob=0.5)
```

```
## [1] 0.875
```

(b)

```
pbinom(q=3,size=6,prob=0.5)-dbinom(x=3,size=6,prob=0.5)
```

```
## [1] 0.34375
```

Problem 12: (R code part)

(a)

```
pexp(q=1,rate=1/2)
```

```
## [1] 0.3934693
```

```
pbinom(q=30,size=100,prob=0.3935)
```

```
## [1] 0.03342515
```

(b)

```
pgamma(q=10,shape=4,scale=2,lower.tail=F)
```

```
## [1] 0.2650259
```

Problem 13: (R code part) (a)

```r
dhyper(x=0,m=3,n=25-3,k=3)
```

```
## [1] 0.6695652
```

(b)

```r
dhyper(x=1,m=1,n=25-1,k=3)
```

```
## [1] 0.12
```

Problem 14: (R code part) (a)

```r
dpois(x=0,lambda=7.5)
```

```
## [1] 0.0005530844
```

Problem 15: (R code part) (a)

```r
pnorm(10150,10000,100,lower.tail=F)
```

```
## [1] 0.0668072
```

(b)

```r
pnorm(10200,10000,100)-pnorm(9800,10000,100)
```

```
## [1] 0.9544997
```

Problem 16: (R code part)

```r
pbeta(q=.8,shape1=3,shape2=2,lower.tail=F)
```

```
## [1] 0.1808
```