Tutorial 1

AID-521: Mathematics for Data Science

Abhishek Samantray, Vimal Kumar

August 7, 2021

Qn. 1.

A traveller in India is preparing an itinerary for a visit to six states. The distance traveled, and hence the cost of the trip, will depend on the order or sequence in which he plans his route.

(a) How many different itineraries (and trip costs) are possible?

(b) Tamil Nadu and Karnataka are two of the states that he plans to visit. The traveller randomly selects one of the possible itineraries. What is the probability that he will visit Karnataka before Tamil Nadu?

Ans/Sol.

(a) 6!

(b) 0.5

Qn. 2.

A car is produced by five production Lines in a manufacturing operation. The cars are shipped to suppliers in 100-unit lots (all units in a lot come from the same Line). Most buyers of the car test only a small number of cars before deciding to accept or reject lots of incoming cars. All five production Lines produce cars at the same rate and normally produce only 1% defective cars, which are dispersed randomly in the output. Unfortunately, production Line 1 suffered mechanical difficulty and produced 5% defectives during the month of March. The manufacturer learned about this after the cars had been shipped. A customer received a lot produced in March and tested three cars – one failed.

(a) What is the probability that the lot was produced on Line 1?

(b) What is the probability that the lot came from one of the four other Lines?

Ans/Sol.

Let A denote the event that a car (and all other cars in the same lot) was produced on Line 1.

The unconditional probability of A is $P(A) = 0.2$, since all production Lines produce cars at the same rate. So, without any additional information, it is equally likely for any Line to

have produced a given car.

In March, there was some extra trouble at Line 1. The customer has received a lot after this event has occurred in March. The customer tests only three cars.

Let B denote the event that exactly 1 car out of 3 cars tested by the customer is defective. (You may imagine this analogous to the event of exactly one heads in three sequential coin tosses.)

Clearly, the probability of B can be affected by the production Line, since Line 1 had higher chances of defect during March. (Conditional Probability!)

If all cars were produced at Line 1, what would be the probability of B?
$P(B|A) =$ (no. of ways to choose 1 defective car out of 3 cars) $\times$ (probability that the 1 chosen car is defective) $\times$ (probability that the other 2 cars are good) $= \binom{3}{1} \times 0.05 \times (0.95)^2 = 0.135$

The rate of producing defective cars is the same in all other Lines. If all cars in the lot of the customer were produced in any Line other than Line 1, what would be the probability of B?
$P(B|A^c) = \binom{3}{1} \times 0.01 \times (0.99)^2 = 0.029$

(a) The manufacturer is already informed about the trouble during March, and is now trying to infer the likelihood that *the lot received by the customer* was *produced on Line 1.*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Hint: Use total probability rule for the denominator, and multiplication law of conditional probability for the numerator. (Bayes' Theorem!)
Ans: 0.538

(b) $P(A^c|B) = 1 - P(A|B) = 0.462$

Qn. 3.

A marketing agent can contact either one or two customers per week with probability 2/5 and 3/5, respectively. Each contact usually results in either no sale or a Rs. 50000 sale, with the probabilities 0.8 and 0.2, respectively.

(a) Write the probability distribution for weekly sales.

(b) Find the mean and standard deviation of the weekly sales.

Ans/Sol.

Let X denote the r.v. for the number of contacts with customers made in a week. Let Y denote the r.v. for the amount, in Rs., of sale done with each (contacted) customer. We can see that both X and Y are discrete random variables.

Let Z be the random variable for weekly sales. Clearly, $Z = X \cdot Y$, and so Z can take values from {range of X} $\times$ {range of Y} $= \{1, 2\} \times \{0, 50000\} = \{0, 50000, 100000\}$

(a) For the probability distribution of Z, we need to associate each value of Z with the probability of such an event.

$P(Z=0) = P(X=1)\cdot P(Y=0) + P(X=2)\cdot[P(Y=0)]^2 = \frac{2}{5}\times 0.8 + \frac{3}{5}\times 0.8^2 = 0.704$

$P(Z=50000) = P(X=1)\cdot P(Y=50000) + P(X=2)\cdot[2\cdot P(Y=0)\cdot P(Y=50000)] = \frac{2}{5}\times 0.2 + \frac{3}{5}\times[2\times 0.8\times 0.2] = 0.272$

$P(Z=100000) = P(X=2)\cdot[P(Y=100000)]^2 = \frac{3}{5}\times 0.2^2 = 0.024$

See that the sum of all probabilities of Z is 1.

(b) $E[Z] = \sum_{z\in Z} z\cdot p(z) = 0\times 0.704 + 50000\times 0.272 + 100000\times 0.024 = 16000$ (rupees)

$Var(Z) = E[(Z-16000)^2] = E[Z^2] - 16000^2$
$= 0^2\times 0.704 + 50000^2\times 0.272 + 100000^2\times 0.024 - 16000^2 = 664000000$
$\implies \sigma(Z) = \sqrt{664000000} = 25768.197$

Qn. 4.

The total hours of the usage of a machine per week is random, and has the probability density function given by

$$f(t) = \begin{cases} \frac{3}{64}t^2(4-t), & 0\le t\le 4, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Calculate the mean and variance of weekly usage of the machine (in hours).

(b) The machine costs Rs. 200 per hour. You are asked to provide an expected value and variance of the weekly cost of the machine.

(c) How often do you expect the weekly cost to exceed Rs. 600?

Ans/Sol.

Let X denote the (continuous) r.v. for the weekly usage of the machine.

(a) $E[X] = \int_{-\infty}^{\infty} x\cdot f(x)\,dx = \int_0^4 x\cdot\frac{3}{64}x^2(4-x)\,dx = \frac{3}{64}\cdot(x^4 - \frac{x^5}{5})\Big|_0^4 = 2.4$ hours

$Var(X) = E(X^2) - 2.4^2 = \int_{-\infty}^{\infty} x^2\cdot f(x)\,dx = \cdots = 0.64$

(b) The r.v. for weekly cost is $200\cdot X$. Let us call this Y.
Hence, required expected value $E[Y] = E[200\cdot X] = 200\times 2.4$ rupees, and
the required variance $Var(Y) = Var(200\cdot X) = 200^2\times 0.64 = 25600$

(c) The required value is $P(Y>600) = P(3\cdot X>600) = P(X>3)$
$= \int_3^{\infty} f(x)\,dx = \int_3^4 f(x)\,dx = 0.2616.$

In other words, we expect that the weekly cost would exceed Rs. 600 in 26.16% of the weeks.