M.Sc. in ‚Transportation Systems‘

Applied Statistics in Transport

Exercises: Distributions

---

1. Given the probability density function $f(x) = \frac{2x+1}{25}$ $(x = 0,1,2,3,4)$; determine the following probabilities:

   a) $\mathrm{P(X = 4)}$

   b) $\mathrm{P(X \leq 1)}$

   c) $\mathrm{P(2 \leq X < 4)}$

   d) $\mathrm{P(X > -10)}$

Solution:

a) $P(X = 4) = \frac{2*4+1}{25} = \frac{9}{25}$

b) $P(X \leq 1) = \frac{2*0+1}{25} + \frac{2*1+1}{25} = \frac{4}{25}$

c) $P(2 \leq \mathrm{X} < 4) = \frac{2*2+1}{25} + \frac{2*3+1}{25} = \frac{12}{25}$

d) $\mathrm{P(X > -10) = 1}$

2. The length of train parts are measured to the nearest tenth of a millimetre. The lengths are uniformly distributed, with values at every tenth of a millimetre starting at 590.0 and continuing through 590.9. Determine the probability density function. Compute P(X≤590.5). Determine the mean and the variance of length.

Solution:

Pdf: f(x)=0.1, $P(X \leq x) = \sum_{x_i \leq x} p(x_i) = P(X \leq 5) = \sum_{x_i \leq 5} 0.1 = 0.6$

$$\mu = \sum_{\mathrm{x}} \mathrm{x_i} * \mathrm{p(x_i)}$$
$$= 590.0 * 0.1 + 590.1 * 0.1 + 590.2 * 0.1 + 590.3 * 0.1 + 590.4 * 0.1 + 590.5 * 0.1 + 590.6 * 0.1 + 590.7 * 0.1 + 590.8 * 0.1 + 590.9 * 0.1 = 590.45$$

$$\sigma^2 = V(X) = E(x-\mu)^2 = \sum (x_i - \mu)^2 * p(x_i) = \sum x_i^2 p(x_i) - \mu^2$$

$$\sigma^2 = (590.0^2 * 0.1 + 590.1^2 * 0.1 + 590.2^2 * 0.1 + 590.3^2 * 0.1 + 590.4^2 * 0.1 + 5 * 0.1 + 590.6^2 * 0.1 + 590.7^2 * 0.1 + 590.8^2 * 0.1 + 590.9^2 * 0.1) - 590.45^2 = 0.0825$$

(590.0^2*0.1+590.1^2*0.1+590.2^2*0.1+590.3^2*0.1+590.4^2*0.1+590.5^2*0.1+590.6^2*0.1 +590.7^2*0.1+590.8^2*0.1+590.9^2*0.1)- 590.45^2

Or simplified:

The expected value of a random value Y is Y= a*X, with X being a random value and a a constant: μ=E(Y) =a*E(X).

For the variance: Var(Y) = a^2 * Var(X)

Let X be the length of train parts, then Y = 10*X has a discrete uniform distribution in 5900 to 5909.

$\mu = E(Y) = \frac{b+a}{2} = \frac{5909+5900}{2} = 5904.5$

E(X) = E(Y)/10 = 590.45

Var(Y) = $((5909-5900+1)^2 -1)/12$ = 99/12 = 8.25

$\sigma^2 = \frac{(b-a+1)^2-1}{12}$

Var(X) = Var(Y)/100 = 0.0825

3. There is a chance that a bit transmitted through a digital transmission channel is incorrectly received. Let X equal the number of incorrectly received bits in the next four transmitted bits. The possible values for X are {0,1,2,3,4}. Suppose that the probabilities are

$$P(X=0)=0.65; P(X=1)=0.29; P(X=2)=0.04; P(X=3)=0.0036; P(X=4)=0.0001$$

Determine the mean and the variance for the probability distribution.

Solution:

$$\mu = 0*f(0)+1*f(1)+2*f(2)+3*f(3)+4*f(4)$$
$$= 0*0.65+1*0.29+2*0.04+3*0.0036+4*0.0001 = 0.3812$$

$$\sigma^2 = \sum (x_i-\mu)^2 * p(x_i) = \sum x_i^2 p(x_i) - \mu^2 = 0.65*0^2+0.29*1^2+0.04*2^2+0.0036*3^2+0.0001*4^2-0.3812^2 = 0.3387, \sigma = 0.5820$$

4. Binomial distribution: Determine the probability to roll exactly one 6 with 8 dice.

Solution:

A=the 6 occurs; $\overline{A}$=any other number occurs ({1,2,3,4,5}) occurs; p=1/6; q=5/6; n=8; x=1

$$p(X=1|n=8) = \binom{8}{1} * \left(\frac{1}{6}\right)^1 * \left(\frac{5}{6}\right)^7 = 0.372$$

5. Binomial: From income and age statistics it is known that 2 percent of the inhabitants in city A are retired persons with a monthly income of less than 800 Euro. What is the probability that at least 5 persons from this group are represented in a random sample of 200 persons?

Solution:

Binomial Distribution, event A – a retired person is in the sample, p=0.02, we look for

$p(X \geq 5|n=200) = 1-P(X \leq 4) = 1-P(X<5)$

$$p(X \geq 5|n=200) = 1-P(X \leq 4)$$
$$= \binom{200}{0} 0{,}02^0 0{,}98^{200} + \binom{200}{1} 0{,}02^1 0{,}98^{199} + \binom{200}{2} 0{,}02^2 0{,}98^{198}$$
$$+ \binom{200}{3} 0{,}02^3 0{,}98^{197} + \binom{200}{4} 0{,}02^4 0{,}98^{196} = 1-0.6288 = 0.3712$$

With dbinom(0,200,0.02)+ dbinom(1,200,0.02)+ dbinom(2,200,0.02)+ dbinom(3,200,0.02) + dbinom(4,200,0.02)= 0.6288

Or: 1-pbinom(4,200,0.02)

6. A batch of parts contains 100 parts from a local supplier of tubing and 200 parts from a supplier of tubing in the next state. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier? What is the probability that two or more parts in the sample are from the local supplier? What is the probability that at least one part in the sample is from the local supplier?

Solution:

phyper(q, m, n, k), dhyper(x, m, n, k)

With

x, q vector of quantiles representing the number of white balls drawn without replacement from an urn which contains both black and white balls.

m the number of white balls in the urn.

n the number of black balls in the urn.

k the number of balls drawn from the urn.

X = the number of parts in the sample from the local supplier, it has a hypergeometric distribution, we ask for P(X=4)

$P(X=4)=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=\frac{\binom{100}{4}\binom{300-100}{4-4}}{\binom{300}{4}}=dhyper(4{,}100{,}200{,}4)=0.0119$

$P(X\geq 2)=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=\frac{\binom{100}{2}\binom{300-100}{4-2}}{\binom{300}{4}}+\frac{\binom{100}{3}\binom{300-100}{3-2}}{\binom{300}{4}}+\frac{\binom{100}{4}\binom{300-100}{4-2}}{\binom{300}{4}}=phyper(4{,}100{,}200{,}4)-phyper(1{,}100{,}200{,}4)=dhyper(4{,}100{,}200{,}4)+dhyper(3{,}100{,}200{,}4)+dhyper(2{,}100{,}200{,}4)=0.4074$

$$P(X\geq 1)=1-P(X<1)=1-P(X=0)=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=1-\frac{\binom{100}{0}\binom{300-100}{4-0}}{\binom{300}{4}}$$
$$=1-dhyper(0{,}100{,}200{,}4)=0.8045$$

7. Queuing space: A maximum of 20 vehicles can stand in the queuing space between a railway crossing gate and the next intersection. The average number of arriving vehicles is 150 vehicles per hour. The gate is closed for a maximum of 4 minutes.
   a) What is the probability that the vehicles jam up to the next intersection?
   b) Due to changes in the timetable the closing time of the gate goes up from 4 to 6 minutes. What is now the probability that the vehicles jam up to the next intersection?

Solution:

a)
$$\lambda=\frac{150veh.}{60min}*4min=\frac{150*4}{60}=10veh./4min.$$
$$P(X>20)=1-P(X\leq 20)=1-ppois(20{,}10)=1-0.9984$$
$$=1-sum\big(dpois(0{:}\,20{,}10)\big)=0.0016$$

With a probability of around 0.2 percent more than 20 vehicles are congesting between the railway crossing gate and the next intersection and this hindering the traffic at the intersection.

b)
$$\lambda=\mathrm{m}=\frac{150veh.}{60min}*6min=\frac{150*6}{60}=15veh./6min.$$

$$P(X > 20) = 1 - P(X \leq 20) = 1 - ppois(20{,}15) = 1 - 0.9170$$
$$= 1 - sum\big(dpois(0{:}\,20{,}15)\big) = 0.083$$

With a probability of around 8 percent more than 20 vehicles are congesting between the railway crossing gate and the next intersection and this hindering the traffic at the intersection. Measures to avoid this congestion should be implemented asap.

8. Determine the probability to get three to six times the head when throwing an ideal coin 10 times. Use (a) the binomial distribution and (b) the normal distribution:

Solution:

(a) $P(X = 3) = \binom{n}{x} * \ p^x * (1-p)^{n-x} = \binom{10}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^7 = \frac{15}{128} = 0.1172$

$P(X = 4) = \binom{10}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6 = \frac{105}{512} = 0.2051$, $P(X = 5) = \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{63}{256} = 0.2461$

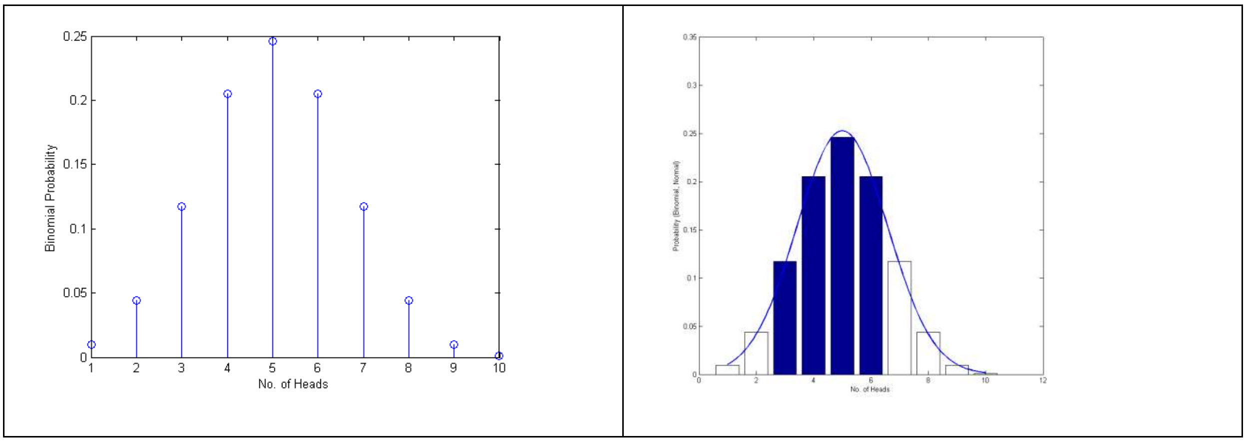
$P(X = 6) = \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4 = \frac{105}{512} = 0.2051,$

$P(3 \leq X \leq 6) = 0.1172 + 0.2051 + 0.2461 + 0.2051 = 0.7735$

sum(dbinom(3:6,10,0.5))= 0.7734

pbinom(6,10,0.5)- pbinom(2,10,0.5)= 0.7734

The probability distribution of numbers of heads when the coin is thrown 10 times is shown in the following diagrams. In the right diagram the data is handled like continuous data. The probability we are interested in is the sum of the area of the dark columns. We can approximate this area with the normal distribution.



(b) Three to six times discrete data can be interpreted as 2.5-6.5 continuous data; from the binomial distribution we have the mean and the variance:

$\mu = n * p = 10 * 0.5 = 5;\ \sigma = \sqrt{n * p * q} = \sqrt{10 * 0.5 * 0.5} = 1.58,$

$\mathrm{pnorm}(2.5{,}5{,}1.58) = \ 0.0568, \mathrm{pnorm}(6.5{,}5{,}1.58) = \ 0.8288;\ 0.8288 - 0.0568 = 0.7720$

Or with tables:

Standardisation: $z_{2.5} = \frac{x-\mu}{\sigma} = \frac{2.5-5}{1.58} = -1.58, z_{6.5} = \frac{6.5-5}{1.58} = 0.95$

We look for the cdf(6.5) minus the cdf(2.5): $z_{6.5} - z_{2.5} = 0.8289 - 0.0571 = 0.7718$

9. Mont 4.3, General continuous distributions: Suppose that $f(x) = x/8$ for 3<x<5. Determine the following probabilities:

a) $P(X < 4)$
b) $P(X > 3.5)$
c) $P(4 < X < 5)$
d) $P(X < 4.5)$
e) $P(X < 3.5 \; or \; X > 4.5)$

Solution:

$P(a \le X \le b) = \int_a^b f(x)dx$

a) $P(X < 4) = \int_3^4 x/8dx = \frac{x^2}{16}\Big|_3^4 = 1 - \frac{9}{16} = 0.4375$

b) $P(X > 3.5) = \int_{3.5}^5 x/8dx = \frac{x^2}{16}\Big|_{3.5}^5 = \frac{25}{16} - \frac{3.5*3.5}{16} = 0.7969$

c) $P(4 < X < 5) = \int_4^5 x/8dx = \frac{x^2}{16}\Big|_4^5 = \frac{25}{16} - 1 = 0.5625$

d) $P(X < 4.5) = \int_3^{4.5} x/8dx = \frac{x^2}{16}\Big|_3^{4.5} = \frac{4.5*4.5}{16} - \frac{9}{16} = 0.7031$

e) $P(X < 3.5 \; or \; X > 4.5) = \int_3^{3.5} x/8dx + \int_{4.5}^5 x/8dx = \frac{x^2}{16}\Big|_3^{3.5} + \frac{x^2}{16}\Big|_{4.5}^5 = \frac{3.5*3.5}{16} - \frac{9}{16} + \frac{25}{16} - \frac{4.5*4.5}{16} = 0.5$

10. Suppose that $f(x) = 0.125x$ for 0<x<4. Determine the mean and variance of X:

Solution:

$\mu = \int_{-\infty}^{\infty} xf(x)dx = \int_0^4 x0.125xdx = \frac{0.125}{3} * 4^3 = 2.6667$

$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = \int_0^4 x^2 0.125xdx - 2.6667^2 = 0.8887$

(4^4)/4*0.125-2.6667^2

11. For two alternative routes the following means μ and standard deviations σ of the travel time x are known:

Route A through the city centre: $\mu_A = 27\ min, \sigma_A = 5\ min$

Route B using a tangent: $\mu_B = 29\ min, \sigma_B = 2\ min$

For which route the risk to come late to an important date is lower, if the departure time is (Case 1) 28 minutes and (Case 2) 32 minutes before the date?

Remark: We assume the travel times to be normally distributed.

Solution: The risk to be late corresponds to the probability that the travel time is higher than the remaining time:

Case 1: P(X>28); Case 2: P(X>32)

Standardisation:

Case 1: Possible Travel time = 28min

Route A: $P(X > 28) = 1 - P(X \leq 28) \rightarrow\ z_A = 1 - F\left(\frac{x_A - \mu}{\sigma}\right) = 1 - F\left(\frac{28-27}{5}\right) = 1 - F(0.2) = 1 - 0.5793 = 0.4207$ (=F(-0.2))

Route B: $P(X > 28) = 1 - P(X \leq 28) \rightarrow\ z_B = 1 - F\left(\frac{x_B - \mu}{\sigma}\right) = 1 - F\left(\frac{28-29}{2}\right) = 1 - F(-0.5) = 1 - 0.3085 = 0.6915$ (=F(0.5))

In this case the risk to come late is higher when taking the tangential route B.

Case 2: Possible Travel time = 32min

Route A: $P(X > 32) = 1 - P(X \leq 32) \rightarrow\ z_A = 1 - F\left(\frac{x_A - \mu}{\sigma}\right) = 1 - F\left(\frac{32-27}{5}\right) = 1 - F(1) = 1 - 0.8413 = 0.1587$

Route B: $P(X > 32) = 1 - P(X \leq 32) \rightarrow\ z_B = 1 - F\left(\frac{x_B - \mu}{\sigma}\right) = 1 - F\left(\frac{32-29}{2}\right) = 1 - F(1.5) = 1 - 0.9332 = 0.0668$

In this case the risk to come late is higher when taking the inner city route A.