BN001, BN009, BN012, BN117, BN121, BN903

Statistics and Probability

Worksheet – Random Variables

Question 1

A factory is producing components, of which 1.9% are defective. They are packed in boxes, each containing 15 components. Using the Poisson approximation to the Binomial distribution,

a) Calculate the probability that a box has 2 defective components.

Solution

Working this as a straightforward binomial question:

The components are packed in boxes of 15, so $n = 15$.

For each component, the probability it is defective, based on the proportion that are known to be defective, is 1.9/100 = 0.019.

$P[X = 2] = {}^{n}C_{r}\, p^{r}(1-p)^{n-r} = {}^{15}C_2 \times 0.019^2 \times 0.981^{13} = 105 \times \text{etc} =$ 0.0295.

To apply the Poisson distribution, we require p is small.

Here $\mu = np = 15\text{x}0.019 = 0.285$.

Then the Poisson distribution gives

$P[X = 2] = e^{-0.285}(0.285)^2/2! = 0.0305$.

This is quite close.

b) It is estimated that a customer receiving a box of components will dismiss one defective but will complain if the box contains 2 or more defectives. Calculate the probability of this happening.

Solution

Since each event $X = r$ is a distinct and independent event, we can say that $P[X \geq 2] = P[X = 2] + P[X = 3] + P[X = 4] =$

We have $P[X = 2] = 0.0305$.

Continue on calculating the probabilities for 3, 4 etc.

$P[X = 3] = e^{-0.285}(0.285)^3/3! = 0.0029$.

$P[X = 4] = e^{-0.285}(0.285)^4/4! = 0.0002$.

$P[X = 5] < 10^{-4}$, so stop here.

The total is then 0.0336.

Alternatively this could have been calculated using

$P[X \geq 2] = 1 - P[X < 2] = 1 - (P[X = 0] + P[X = 1])$

c) If 20 of these boxes are sold, identify the distribution for the random variable of the number of potential complaints.

Solution

Since there is a fixed probability that one box gets a complaint, 0.0336, and this is repeated with the 20 boxes sent, this is a binomial distribution.

Question 2

A multiple choice exam has 10 questions, each one with a choice of four answers, only one of which is correct. Let X be the random variable of the number of correct answers for a student picking their answers at random.

a) Identify the distribution governing variable X and therefore the expected value of correct questions for this student. Find the standard deviation for X.

Solution

The distribution is the binomial distribution with $n = 10$ and $p = ¼$.

The expected value is 10x¼ = 2.5.

The variance is $\sigma^2 = np(1 - p) = 10 \text{ x } ¼ \text{ x } ¾$.

Then $\sigma = ¼\sqrt{30} = 1.37$ approx.

b) Calculate the probability of the student getting no questions correct.

Solution

This is simply $¾^{10} = 0.0563$.

c) Calculate the probability of the student getting more than 3 questions correct, in other words a pass mark.

Solution

The event '$X > 3$' is the event of a pass.

The event '$X = 0$ or $X = 1$ or $X = 2$ or $X = 3$' is a fail.

They are the exact converse; in the context of the question.

So P['Pass'] = 1 – P['Fail'],

P[‘Fail’] = $P[X = 0$ or $X = 1$ or $X = 2$ or $X = 3] =$

$= P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3]$.

We have already worked out the first two, and:

$P[X = 2] = {}^{10}C_2 \times 0.25^2 \times 0.75^8 = 0.2816$, $P[X = 3] = 0.2503$, and so

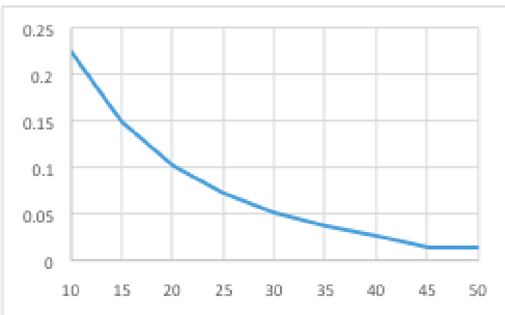
P[‘Fail’] = 0.7759.

$P[X > 3] = 1 – 0.7759 = 0.2241$.

d) Consider the more general case of n questions, where each question has k possible answers with only one correct, so $pk = 1$. Create a graph of the function $f(n) = P[X > 0.4n]$ for $k = 3$ to 6, that is, the probability of passing against the number of question for a given number of answers. Describe what happens as n increases, explaining why. Repeat this for $k = 2$.

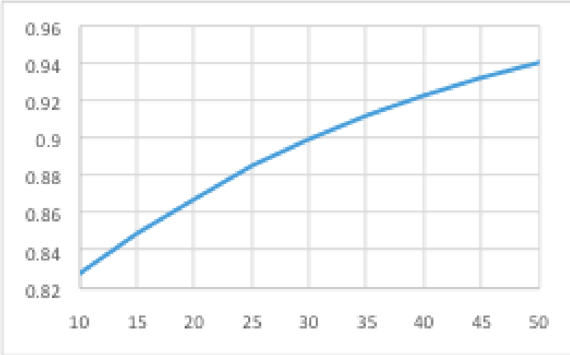
Solution

Here is the graph for $k = 4$, so $p = ¼$. The horizontal axis is n.



Clearly, the more questions, the less likely it is for the student to get to the $0.4n$ pass mark. This is because the expected value is $np = n/4$ with a standard deviation of $\sigma = ¼\sqrt{(3n)}$, so the ‘distance’ of $0.25n$ and $0.4n$ widens.

Here is the graph for $k = 2$, so $p = ½$. The horizontal axis is n.



in this case, the more questions, the more likely it is for the student to get to the $0.4n$ pass mark. This is because the expected value ($np = n/2$) is above the pass mark, with the 'distance' of $0.5n$ and $0.4n$ widening.

Question 3

A guest at a gambling club plays a game for 10€, where a player asks them to choose a number 1 to 6, then rolls three dice. If the chosen number does not come up, the guest does not win. If the number comes up once, the guest wins 10€, twice wins 20€ and three times wins 30€.

a) Identify the exact distribution for the number of times the number chosen by the guest comes up.

Solution

Let R be the random variable of the number of times the chosen number comes up; the distribution is the binomial distribution with $n = 3$ and $p = 1/6$.

$P[R = r] = {}^3C_r\,(1/6)^r(5/6)^{3-r}$.

The expected value is 3/6 = ½.

b) Using the expected value as a measure, calculate whether the game favours the guest or the player. Explain why the expected value is a good measure for this.

Solution

The expected value of the mount of money gained X will be used. The wins or loses over time will tend towards this mean amount as more games are played.

The net gain after a game is

$E[X] = -10P[R = 0] + 10P[R = 1] + 20P[R = 2] + 30P[R = 3]$.

This is simply every outcome with its probability.

Using the distribution is the binomial distribution with $n = 3$ and $p = 1/6$;

$E[X] = -10(5/6)^3 + 10x3x(1/6)(5/6)^2 + 20x3x(1/6)^2(5/6) + 30x(1/6)^3 =$

$= -10x5^3/6^3 + 10x3x5/6^3 + 20x3x5/6^3 + 30/6^3$.

The top line is –1250 + 150 + 300 + 30 = 480 – 1250 = –770.

Therefore $E[X] = -3.565$.

The balance is very much in favour of the game rather than the player.

Question 4

Show that for a small value of p, the Binomial distribution becomes the Poisson as n becomes very large.

Solution

The first step is the approximation of $n!$ as n gets very large:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n .$$

Use this in the equation for the quantity ${}^nC_r$:

$${}^nC_r = \frac{n!}{(n-r)!r!} \cong \sqrt{\frac{n}{n-r}}\left(\frac{n}{n-r}\right)^n \frac{e^{-r}(n-r)^r}{r!} .$$

This is then ${}^nC_r = \dfrac{n!}{(n-r)!r!} \cong \sqrt{\dfrac{n}{n-r}}\,\dfrac{1}{(1-\frac{r}{n})^n}\,\dfrac{e^{-r}(n-r)^r}{r!}$.

Letting n get very large means $\lim_{n\to\infty}\left(1-\dfrac{r}{n}\right)^n = e^{-r}$.

So overall this will be ${}^nC_r \sim (n-r)^r/r!$

Now write the expression for the binomial PMF; the assumption that p is small means we can treat $\mu = np$ as a number like r; not large.

$$P[N=r] \cong \frac{(n-r)^r}{r!}\left(\frac{\mu}{n}\right)^r\left(1-\frac{\mu}{n}\right)^{n-r} = \frac{\mu^r}{r!}\left(\frac{n-r}{n}\right)^r \frac{1}{\left(1-\frac{\mu}{n}\right)^r}\left(1-\frac{\mu}{n}\right)^n .$$

Bring together the terms with r in the power:

$$P[N=r] \cong \frac{\mu^r}{r!}\left(\frac{n-r}{n-\mu}\right)^r\left(1-\frac{\mu}{n}\right)^n .$$

Letting n become very large means:

$$\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n = e^{-\mu}, \ \lim_{n\to\infty}\left(\frac{n-r}{n-\mu}\right)^r = 1 .$$

Therefore $P[N=r] = e^{-\mu}\mu^r/r!$ as required.

Question 5

A company operates a tollgate on an Autoroute in the Pyrénées. It has been established that for the month of October, cars arrive at this booth at a rate of one every 3 minutes.

a) Calculate the probability that a car arrives within 6 minutes.

b) Calculate the probability that 3 cars arrive in a ten-minute period for this tollbooth.

c) For another tollbooth, the rate is one car every 4 minutes. Calculate the probability that another car arrives within 8 minutes.

d) Compare the answers for part (a) and (b), explaining any similarities/differences.

Question 6

A communications device relies on a power source that experiences a power surge once every 4 days.

a) Calculate the probability that a power surge occurs within the next 2 days, and then for the next 6 days.

Solution

The rate information is '1 power surge every 4 days', so this is a rate:

λ = ¼ = 0.25 surges/day.

The period of time is 2 days. So if T is the random variable of the amount of time till the next event, then

$$\lambda t = ¼ \text{ x } 2 = ½ = 0.5.$$

$$P[T < 2] = 1 - e^{-0.25\text{x}2} = 1 - e^{-0.5} = 1 - 0.6065 = 0.3935.$$
$$P[T < 6] = 1 - e^{-0.25\text{x}6} = 1 - e^{-1.5} = 1 - 0.2231 = 0.7769.$$

b) It is suspected that two power surges within one day will impair the functions of the device. Calculate the probability of this happening.

Solution

This is now the Poisson distribution, with parameter $\mu = \lambda T$.

This is: $\mu = ¼ \text{ x } 1 = 0.25$.

$$P[X = 2] = e^{-0.25}(0.25)^2/2! = 0.7788\text{x}0.0625/2 = 0.0243.$$

Question 7

Two players are in a game of Russian roulette. The revolver has six chambers and one live round is being used. After each play, the cylinder is rotated so that the next play is an independent event. The game ends once the revolver has fired and that player who fired the revolver loses. Let variable N be the number of shots fired in a game.

a) Use the laws of probability to write down a distribution for the variable N. Need the possibility of the game continuing endlessly be included?

Solution

Let B be the event that the player does not lose and let L be the event that the player loses the game.

The event that $N = r$ is the event of the sequence $BBBB\ldots L$, where there are $r - 1$ consecutive independent B events.

Let p denote the probability of the revolver not firing, which is 5/6.

Then $P[N = r] = p^{r-1}(1 - p)$. This is the PMF.

So in this instance $P[N = r] = (1/6)(5/6)^{r-1}$.

As r becomes very large, this tends to 0, so there is no need to include the infinite case.

b) Calculate the probability that the same player who starts the game loses. [Hint: recall that for any number k, $2k + 1$ is an odd number and $2k$ is an even number]

Solution

The event that the player who starts loses is the event that the game plays out as

L, *BBL*, *BBBBL*, *BBBBBBL*, …, that is, all the odd values of r.

These are all mutually exclusive independent events therefore the sum over any $r = 2k + 1$, so

$$P[N \text{ is odd}] = (1-p)\sum_{k=0}^{\infty} p^{2k} = (1-p)\sum_{k=0}^{\infty}\left(p^2\right)^k = \frac{1-p}{1-p^2} = \frac{1}{1+p}.$$

For $p = 5/6$ this is 5/11, just over 0.5.

c) Identify the expected value and standard deviation for the variable N.

Since $P[N = r] = p^{r-1}(1 - p)$ with $p = 5/6$.

The expected value of N is $E[N] = \sum_{r=1}^{\infty} r(1-p)p^{r-1} = (1-p)\sum_{r=1}^{\infty} rp^{r-1}$.

If $f(x)=\sum_{r=0}^{\infty} x^r = \frac{1}{1-x}$, then $f'(x) = \sum_{r=1}^{\infty} r x^{r-1} = \frac{1}{(1-x)^2}$.

The sum may be regarded as starting at r = 1.

Therefore $\sum_{r=1}^{\infty} r p^{r-1} = \frac{1}{(1-p)^2}$.

The expected value is then $E[N]$ = $(1-p)\sum_{r=1}^{\infty} r p^{r-1} = \frac{1-p}{(1-p)^2} = \frac{1}{1-p}$.

With p = 5/6 this gives $E[N]$ = 6.

The second derivative is $f''(x)$ = $\sum_{r=1}^{\infty} r(r-1)x^{r-2} = \frac{2}{(1-x)^3}$.

Write this as $\sum_{r=0}^{\infty} r(r+1)\,p^{r-1} = \frac{2}{(1-p)^3}$.

Therefore $(1-p)\sum_{r=1}^{\infty} r^2 p^{r-1} + (1-p)\sum_{r=1}^{\infty} r p^{r-1} = \frac{2}{(1-p)^2}$.

So $E[N^2] = 2/(1-p)^2 - E[N] = 2\text{x}6^2 - 6 = 66$.

Using these expressions in the definition of the variance gives

$\text{Var}[N] = 66 - 6^2 = 30$, so the standard deviation is $\sigma = \sqrt{30}$.

d) Redo the calculation for part (b) for the case of two live rounds in the cylinder.

Question 7

The following questions use the normal approximation of the binomial distribution for large n.

a) Write down equations giving the values of the mean and standard deviation for this approximation in terms of the parameters n and p.

Solution

The values for the mean and standard deviation are those for the binomial distribution itself. See notes.

b) Assume that a pregnancy results in a boy or girl with equal probability. A large maternity hospital handles 75 births every week and a small one handles 25. Without actually calculating the probabilities, decide which hospital is more likely to produce a given proportion of girls above 0.5.

Solution

See notes for a discussion on the approximation for the binomial. The higher value for the number n will mean a closer to the normal approximation. Therefore the lower number of births permits greater variation. The standard deviation for the binomial distribution itself for the case $p = ½$ is $\sqrt{n}/2$.

c) Calculate the probability that the proportion of girls exceeds a given value p, using the normal approximation. Comment on which of the probabilities in part (b) will be closest to this value.

Solution

Again, see notes for a discussion on the approximation for the binomial. Applying the central limit theorem the quantity shown will follow the standard normal distribution:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Replacing the sample mean $\bar{X}$ with the sample proportion $\hat{p}$, and using $\sigma^2 = n\frac{1}{2}(1 - \frac{1}{2})$, we get

$$\frac{\hat{p} - \frac{1}{2}}{\frac{1}{2}} = 2\hat{p} - 1.$$

This will follow the standard normal distribution, with the convergence increasing faster the closer the true proportion is to ½.