

Statistical methods: Homework 6

Cameron McIntyre

October 2018

1 4.2.4

A chromosome mutation linked with colorblindness is known to occur, on the average, once in every ten thousand births.

Answer:

- (a) Approximate the probability that exactly three of the next twenty thousand babies born will have the mutation.

$$\lambda = 20000 * \frac{1}{10000} = 2$$

$$P(X = 3) = \frac{1}{e^\lambda} \frac{\lambda^n}{n!} = \frac{1}{e^2} \frac{(2)^3}{3!} = .1804$$

- (b) How many babies out of the next twenty thousand would have to be born with the mutation to convince you that the one in ten thousand estimate is too low.

$$P(X > n) = 1 - P(X \leq n-1) = 1 - \sum_{k=0}^{n-1} P(X = k) = 1 - \sum_{k=0}^{n-1} \frac{1}{e^\lambda} \frac{\lambda^k}{k!}$$

We go to R to get some values:

```
1 -cumsum(dpois(c(0:10),2))
```

$$P(X \geq 1) = 0.864664$$

$$P(X \geq 2) = 0.5939942$$

$$P(X \geq 3) = 0.3233236$$

$$P(X \geq 4) = 0.1428765$$

$$P(X \geq 5) = 0.05265302$$

$$P(X \geq 6) = 0.01656361$$

$$P(X \geq 7) = 0.00453380$$

$$P(X \geq 8) = 0.00109671$$

$$P(X \geq 9) = 0.00023744$$

$$P(X \geq 10) = 0.0000464$$

$$P(X \geq 11) = 0.0000083$$

If there are 3 or more births with the mutation in 20000. then I would say the estimate needs to be revised because that should only happen 32% of the time. Also 60% of the time there will be 2 children with the defect. This is more than half the time. This also makes me think the estimate should be revised.

2 4.2.12

Midwestern Skies books ten commuter flights each week. Passenger totals are much the same from week to week, as are the numbers of pieces of luggage that are checked. Listed in the following table are the numbers of bags that were lost during each of the first forty weeks in 2009. Do these figures support the presumption that the number of bags lost by Midwestern during a typical week is a Poisson random variable?

Answer: The average number of bags lost in a given week is 1.525 bags/week. Using R we calculate probabilities

```
> dpois(c(0:10),1.525)
[1] 2.176211e-01 3.318721e-01 2.530525e-01 1.286350e-01 4.904210e-02
[6] 1.495784e-02 3.801784e-03 8.282459e-04 1.578844e-04 2.675263e-05
[11] 4.079776e-06
> sum(dpois(c(0:10),1.525))
[1] 0.9999994
```

$$P(X = 0) = .2176211 \text{ versus } 22.5\% \text{ from the data}$$

$$P(X = 1) = .3318721 \text{ versus } 32.5\% \text{ from the data}$$

$$P(X = 2) = .2530525 \text{ versus } 25\% \text{ from the data}$$

$$P(X = 3) = .1286350 \text{ versus } 12.5\% \text{ from the data}$$

$$P(X = 4) = .04904210 \text{ versus } 5\% \text{ from the data}$$

$$P(X = 5) = .01495784$$

$$P(X = 6) = .003801784$$

$$P(X = 7) = .0008282459$$

$$P(X = 8) = .0001578844$$

$$P(X = 9) = .00002675263$$

$$P(X = 10) = .000004079776$$

The data from the airline is definitely modelled well by a poisson distribution.

3 4.2.28

Fifty spotlights have just been installed in an out-door security system. According to the manufacturer's specifications, these particular lights are expected to burn out at the rate of 1.1 per one hundred hours. What is the expected number of bulbs that will fail to last for at least seventy-five hours?

Answer:

For this scenario $\lambda = 1.1 \frac{\text{bulbs}}{100 \text{ hours}}$ Therefore the probability of a single bulb lasting less than 75 hours is

$$P(H < 75) = (1 - e^{-1.1 \cdot \frac{75}{100}}) = 1 - e^{-1.1 \cdot .75} = .5617$$

$$50 \cdot .5617 = 28.09$$

So, we can estimate that out of the 50 new lights, 28 will have failed.

4 4.3.2

(a) Let $0 < a < b$. Which number is larger?

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \text{ or } \int_{-b}^{-a} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Answer: The two quantities are equal. We rely on the symmetry about 0 of the normal distribution to show this.

Consider only $f(z) = e^{-\frac{z^2}{2}}$. For any number a we can show that $f(-a) = f(a)$.

$$f(-a) = e^{-\frac{(-a)^2}{2}} = e^{-\frac{(-1)^2 a^2}{2}} = e^{-\frac{(1 \cdot a^2)}{2}} = f(a)$$

Therefore,

$$\begin{aligned} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} &= P(a < Z < b) = 1 - (P(Z < a) + P(Z > b)) \\ &= 1 - (P(Z > -a) + P(Z < -b)) \text{ (by symmetry)} \\ &= P(-b < Z < -a) \\ &= \int_{-b}^{-a} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \end{aligned}$$

(b) Let $a > 0$. Which number is larger?

$$\int_a^{a+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \text{ or } \int_{a-\frac{1}{2}}^{a+\frac{1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Answer:

$\int_{a-\frac{1}{2}}^{a+\frac{1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ is the larger quantity.

Consider only $f(z) = e^{-\frac{z^2}{2}}$

Then, $f'(z) = -ze^{-\frac{z^2}{2}}$, Clearly $f'(z) < 0 \forall z > 0$.

Now $(a+1-a) = 1$ and also $(a+\frac{1}{2} - (a-\frac{1}{2})) = 1$ Therefore the integrals are over the same length domain.

But, $f(a+\epsilon) < f(a)$ because $f'(a) < 0$. Therefore the normal function is always less over the interval $[a, a+1]$, compared to $[a-\frac{1}{2}, a+\frac{1}{2}]$.

So we can then conclude

$$\int_a^{a+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} < \int_{a-\frac{1}{2}}^{a+\frac{1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

5 4.3.14

A sell-out crowd of 42,200 is expected at Cleveland's Jacobs Field for next Tuesday's game against the Baltimore Orioles, the last before a long road trip. The ballpark's concession manager is trying to decide how much food to have on hand. Looking at records from games played earlier in the season, she knows that, on the average, 38% of all those in attendance will buy a hot dog. How large an order should she place if she wants to have no more than a 20% chance of demand exceeding supply?

Answer:

Our estimate for μ is $\mu = np = 42200 \cdot .38 = 16036$.

Our estimate for σ is $\sigma = \sqrt{np(1-p)} = 42200 \cdot .38 \cdot .62 = 99.7112$

We need to find the number of hot dogs (**X**) such that the probability of running out does not exceed 20's.

Using R we find the z score for 80's.

```
> qnorm(.8)
[1] 0.8416212
```

We can get our estimate for **X** in the following way.

$$\frac{X - \mu}{\sigma} = \frac{X - 16036}{99.711} = 0.8416212$$

$$X = 16,120$$

Therefore the person running the hot dog stand should order 16,120 hotdogs, if they are aiming to have an 80% chance they don't run out (equivalently 20% chance they do run out).

6 4.3.18

Suppose X_1, X_2, X_3 and X_4 are independent Poisson random variables, each with parameter $\lambda = 3$. Let $S = X_1 + X_2 + X_3 + X_4$.

- (a) Use the Central Limit Theorem to approximate the probability that $13 \leq S \leq 14$. We get our approximations $\mu_s = 4 * 3$ and $\sigma_s = \sqrt{4 * 3} = 3.46$

$$z(S) = \frac{S - \mu_s}{\sigma_s}$$

And $z(14) = .58$, similarly $z(13) = .29$

Now Using the tables $P(.29 < Z < .58) = .1050$

- (b) Calculate the exact probability that $13 \leq S \leq 14$.

We can add Poisson Variable together. So $\mu = 3 + 3 + 3 + 3 = 12$ and

$$\sum_{i=13}^{14} e^{-12} \frac{12^i}{i!} = e^{-12} \frac{12^{13}}{13!} + e^{-12} \frac{12^{14}}{14!} = .1056 + .0905 = .1961$$

We could eliminate some of the difference by adding in the continuity correction from the lectures.

7 4.3.22

A large computer chip manufacturing plant under construction in Westbank is expected to result in an additional fourteen hundred children in the county's public school system once the permanent workforce arrives. Any child with an IQ under 80 or over 135 will require individualized instruction that will cost the city an additional \$1750 per year. How much money should Westbank anticipate spending next year to meet the needs of its new special ed students? Assume that IQ scores are normally distributed with a mean (μ) of 100 and a standard deviation (σ) of 16.

Answer:

We can assume $IQ \sim N(100, 16)$. We need to estimate the number of children with and $IQ > 135$ and others with and $IQ < 80$.

$$P(X > 135) = P\left(\frac{X - \mu}{\sigma} > 2.1875\right) = P(z > 2.1875) = 0.01435302$$

$$P(X < 85) = P\left(\frac{X - \mu}{\sigma} < -1.25\right) = P(z < -1.25) = 0.1056498$$

From this we can estimate that $0.01435302 \cdot 1400 = 20.09$ Kids will be on the high end and $0.1056498 \cdot 1400 = 147.909$ will be on the low end. That would suggest they will be spending approximately $(147.909 + 20.09) * 1750 = 293,858.25$ on special education.

8 4.4.6

Suppose three fair dice are tossed repeatedly. Let the random variable X denote the roll on which a sum of 4 appears for the first time. Use the expression for $FX(t)$ given in Question 4.4.5 to evaluate $P(65 \leq X \leq 75)$.

Answer:

We have to figure out the probability p of the successful event, namely, the sum being equal to 4. This can happen 3 possible ways. Therefore $p = \frac{6}{216}$

Our expression

$$F(X(t)) = P(X \leq t) = 1 - (1 - p)^t$$

Can then be used in the following way:

$$\begin{aligned} P(65 \leq X \leq 75) &= P(X \leq 75) - P(X \leq 64) = \left(1 - \left(1 - \frac{6}{216}\right)^{75}\right) - \left(1 - \left(1 - \frac{6}{216}\right)^{64}\right) \\ &= 0.8791029 - 0.8351857 = 0.04391719 \end{aligned}$$

9 4.4.10

Suppose that the random variables X_1 and X_2 have mgfs $M_{X_1}(t) = \frac{\frac{1}{2}e^t}{1 - (1 - \frac{1}{2})e^t}$ and $M_{X_2}(t) = \frac{\frac{1}{4}e^t}{1 - (1 - \frac{1}{4})e^t}$, respectively. Let $X = X_1 + X_2$. Does X have a geometric distribution? Assume that X_1 and X_2 are independent.

Answer: Using Theorem 3.12.3, if $X = X_1 + X_2 + \dots + X_n$, and each of the X_i are independent with MGFs $M_{X_i}(t)$. Then, $M_X(t) = \prod_1^n M_{X_i}(t)$.

We also know the mgf of the geometric distribution $\frac{pe^t}{1 - (1 - p)e^t}$. Therefore $X_1 \sim \text{geom}(\frac{1}{2})$ and $X_2 \sim \text{geom}(\frac{1}{4})$.

The product of their MGFs is:

$$M_X(t) = M_{X_1}(t) \cdot M_{X_2}(t) = \frac{\frac{1}{2}e^t}{1 - (1 - \frac{1}{2})e^t} \cdot \frac{\frac{1}{4}e^t}{1 - (1 - \frac{1}{4})e^t} = \frac{\frac{1}{8}e^{2t}}{1 - e^t(\frac{3}{4}e^t - \frac{1}{4})}$$

This new MGF looks nothing like a geometric MGF. Therefore $X = X_1 + X_2$ is not geometrically distributed

10 4.5.5

For a negative binomial random variable whose pdf is given by Equation 4.5.1, find $E(X)$ directly by evaluating $\sum_r \binom{k-1}{r-1} p^r (1-p)^{k-r}$. (Hint: Reduce the sum to one involving negative binomial probabilities with parameters $r + 1$ and p .)

Answer:

$$E(K) = \sum_{k=r}^{\infty} k \binom{k-1}{r-1} p^r (1-p)^{k-r} = \sum_{k=r}^{\infty} \frac{r}{p} \frac{k * (k-1)!}{(k-r)! r(r-1)!} p^{r+1} (1-p)^{k-r}$$

$$\frac{r}{p} \sum_{k=r}^{\infty} \binom{k}{r} p^{r+1} (1-p)^{k-r} = \frac{r}{p} * 1 = \frac{r}{p}$$

11 4.6.2

A service contract on a new university computer system provides twenty-four free repair calls from a technician. Suppose the technician is required, on the average, three times a month. What is the average time it will take for the service contract to be fulfilled?

Answer:

Using the expected value property of the gamma distribution the average time a contract will take to be fulfilled is $\frac{r}{\lambda} = \frac{24}{3} = 8$ months.

12 4.6.6

Prove that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. (Hint: Consider $E(Z^2)$, where Z is a standard normal random variable.)

Answer: This question requires a bit of extra machinery. Our strategy is to take the gamma function, assume it has a finite interval, make a variable substitution, then find the square of that integral, transform into polar co-ordinates, integrate over the new bounds and then hope that it equals $\sqrt{\pi}$ - which it does because its a well established result.

So,

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

We let $\alpha = \frac{1}{2}$ substitute $x = u^2$, (then $dx = 2udu$) and we get,

$$\Gamma(\frac{1}{2}) = \int_0^{\infty} u^{2*\frac{1}{2}} e^{-u^2} 2udu = 2 \int_0^{\infty} e^{-u^2} du$$

Lets assume that this integral has value a finite value. Then the product of two identical integrals will have a finite value.

$$\left(\Gamma(\frac{1}{2})\right)^2 = 2 \int_0^{\infty} e^{-u^2} du \cdot 2 \int_0^{\infty} e^{-v^2} dv = 4 \int_0^{\infty} \int_0^{\infty} e^{-(u^2+v^2)} dudv$$

We now change to polar co-ordinates:

$$\left(\Gamma\left(\frac{1}{2}\right)\right)^2 = 4 \int_0^\infty \int_0^{\frac{\pi}{2}} e^{-r^2} r d\theta dr$$

Substitute $z = r^2$, then $dz = 2rdr$

$$\left(\Gamma\left(\frac{1}{2}\right)\right)^2 = 2 \int_0^{\frac{\pi}{2}} \int_0^\infty e^{-z} dz d\theta = 2 \int_0^{\frac{\pi}{2}} [-e^{-z}] \Big|_0^{z=\infty} = 2 \int_0^{\frac{\pi}{2}} -[0-1] d\theta = 2 \left[\frac{\pi}{2} - 0\right] = \pi$$

So,

$$\left(\Gamma\left(\frac{1}{2}\right)\right)^2 = \pi. \leftrightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

13 4.6.9

Differentiate the gamma moment-generating function to verify the formulas for $E(Y)$ and $\text{Var}(Y)$ given in Theorem 4.6.3.

Answer:

Gamma Moment generating function:

$$M_Y(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}$$

Finding the expectation, we differentiate wrt t and evaluate at $t=0$.

$$E[Y] = M'_Y(t) \Big|_{t=0} = -r \left(1 - \frac{t}{\lambda}\right)^{-r-1} * \frac{-1}{\lambda} \Big|_{t=0} = \frac{r}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{r-1} \Big|_{t=0} = \frac{r}{\lambda}$$

So as expected $E[Y] = \frac{r}{\lambda}$.

For $E[Y^2]$ we find $M''_Y(t) \Big|_{t=0}$

$$\frac{d}{dt} \frac{r}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{-(r+1)} \Big|_{t=0} = \frac{r(r+1)}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{-(r+1)} \frac{-1}{\lambda} \Big|_{t=0} = \frac{r(r+1)}{\lambda^2}$$

Then the variance is $\text{Var}[Y] = E[Y^2] - (E[Y])^2$

$$\text{Var}[Y] = E[Y^2] - (E[Y])^2 = \frac{r(r+1)}{\lambda^2} - \frac{r^2}{\lambda^2} = \frac{r}{\lambda^2}$$

PROGRAMMING ASSIGNMENT
RESULTANT GRAPH FOLLOWS LINE Y=X, SUGGESTING
THE SUM OF 3 POISSON VARIABLES IS POISSON DISTRIBUTED,
LINE Y=X SHOWN IN BLUE.

Part A

```

poi1 <- qpois(runif(1000),2)
poi2 <- qpois(runif(1000),2)
poi3 <- qpois(runif(1000),2)

poi_combined = poi1 + poi2 + poi3

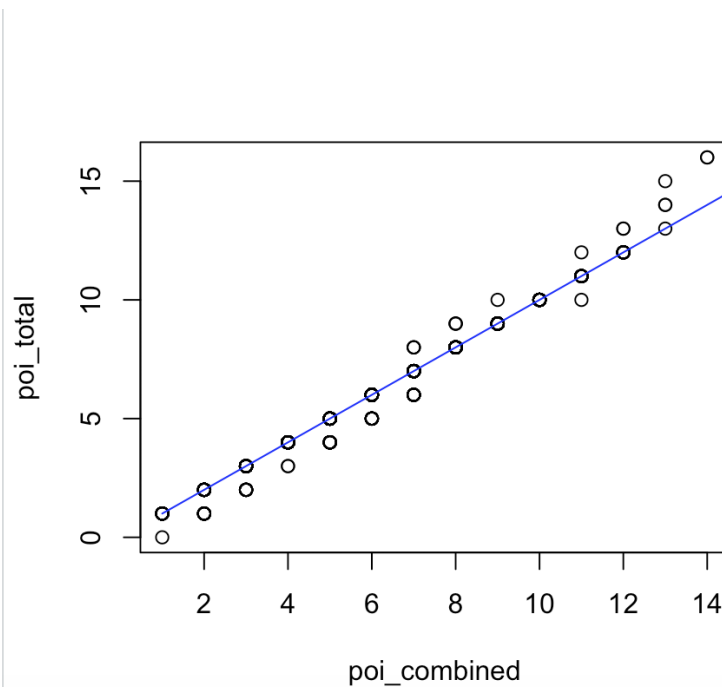
poi_total <- qpois(runif(1000),6)

poi_combined = sort(poi_combined)

poi_total = sort(poi_total)

plot(poi_combined,poi_total)

```



Part B

WE KNOW THAT THE SUM OF THE 3 EXPONENTIAL VARIABLES ARE DISTRIBUTED AS GAMMA WITH PARAMETER $\text{LAMBDA} = \frac{1}{2}$ AND SHAPE PARAMETER 3. WE SHOW THIS GRAPHICALLY BY SUMMING 3 EXPONENTIAL RANDOM VARIABLES AND GRAPHING THEM AGAINST A GAMMA RANDOM VARIABLE. LINE $Y=X$ SHOWN IN BLUE

```

exp1 <- qexp(runif(1000),.5)
exp2 <- qexp(runif(1000),.5)
exp3 <- qexp(runif(1000),.5)

```

```
exp_combined = exp1 + exp2 + exp3

exp_total <- qgamma(runif(1000),shape=3, rate =.5)

exp_combined = sort(exp_combined)

exp_total = sort(exp_total)

plot(exp_combined[51:950],exp_total[51:950], col= 'green')
lines(x= c(1:12),y = c(1:12),col='blue', add= T)
plot(exp_combined[51:950],exp_total[51:950])
```

