problem 1: set-up: you are interested in studying the writing style of a popular Time Magazine contributor, FZ. you collect a simple random sample of his articles and count how many times he uses the word however in each of the articles in your sample, $(x_1, ..., x_n)$. In this set-up, xi is the number of times the word however appeared in the i-th article.

question 1.1 (10 points): define the population of interest, the population quantity of interest, and the sampling units.

Population of interest: all articles from a popular Time Magazine contributor, FZ;
Population quantity of interest: number of times word "however" appear in FZ's article;
Sampling units: aticle;

question 1.2. (10 points): what are potentially useful estimands for studying writing style? (hint: you are interested in comparing FZ writing style to that of other contributors.)

Potential useful estimates for studying the writing style are the length of the article, sentence length, paragraphs indentation in the article, frequency of the word "I" in the article, etc.

question 1.3. (10 points): model: let X_i denote the quantity that captures the number of times the word however appears in the i-th article. let's assume that the quantities $(X_1, ..., X_n)$ are independent and identically distributed (IID) according to a Poisson distribution with unknown parameter λ,

$$p(X_i = x_i \mid \lambda) = Poisson(xi \mid \lambda) \text{ for } i = 1, ..., n.$$

using the 2-by-2 table of what's variable/constant versus what's observed/unknown, declare what's the technical nature (random variable, latent variable, known constant or unknown constant) of the quantities involved the set-up/model above: $X_1, ..., X_n$, $x_1, ..., x_n$, λ and n.

	variable	constants
observed	$X_1, ..., X_n$	$x_1, ..., x_n$ n
not observed		λ

question 1.4. (10 points): write the data generating process for the model above.

for $i = 1...n$ **do**
set λ
sample $x_i \sim$ Poisson(λ); assuming x_i is IID
end for

question 1.5. (10 points): define the likelihood $L(\lambda) = p(\cdot \mid \cdot)$ for this model and set-up at the highest level of abstraction.

$$p(X = x \mid \lambda) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

question 1.6. (10 points): write the likelihood $L(\lambda)$ for a generic sample of n articles, $(x_1, ..., x_n)$.

$L(\lambda) = P(X_1 = x_1 \mid \lambda) \cdot P(X_2 = x_2 \mid \lambda) \cdot P(X_3 = x_3 \mid \lambda)...P(X_n = x_n \mid \lambda)$
$L(\lambda) = \prod_{i=1}^{n} P(X_i = x_i \mid \lambda)$
$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!}$

question 1.7. (10 points): write the log-likelihood $\ell(\lambda)$ for a generic sample of n articles, $(x_1, ..., x_n)$.

$\log(L(\lambda \mid x)) = \ell(\lambda \mid x) = \sum_{i=1}^{n}(-\lambda + x_i \log(\lambda) - \log(x_i!)) = -n \cdot \lambda + \log(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!)$

question 1.8. (10 points): write the log-likelihood $\ell(\lambda)$ for the following specific sample of 7 articles (12, 4, 5, 3, 7, 5, 6).

$n = 7$
$X = (12, 4, 5, 3, 7, 5, 6)$
$\log(L(\lambda \mid x)) = -n \cdot \lambda + \sum_{i=1}^{n} \log(\lambda) x_i - \sum_{i=1}^{n} \log(x_i!)$
$\sum_{i=1}^{n} x_i = 12 + 4 + 5 + 3 + 7 + 5 + 6 = 42$
$\sum_{i=1}^{n} \log(x_i!) = 21.55$
$\log(L(\lambda \mid x)) = -7 \cdot \lambda + 42 \cdot \log(\lambda) - 21.55$

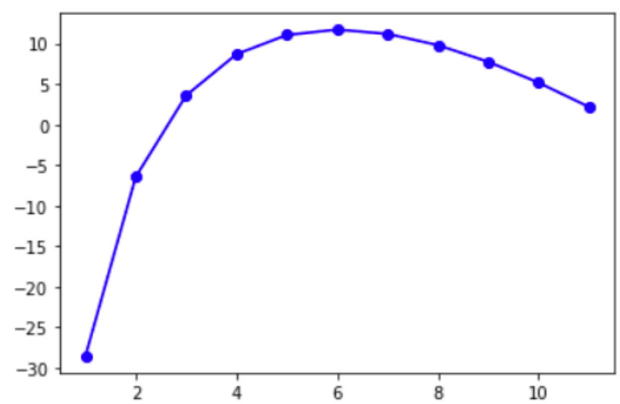
question 1.9. (10 points): plot the log-likelihood $\ell(\lambda)$ in R for the same specific sample of 7 articles (12, 4, 5, 3, 7, 5, 6). What is the maximum value of λ (approximately)?

```python
def poisson_log_likelihood(l):
    res = -7*l+ 42*math.log(l)-21.55
    return res
```

```python
lambda_dict = {}
for i in range(1, 12):
    lambda_dict[i] = poisson_log_likelihood(i)
```

```python
plt.plot(list(lambda_dict.keys()), list(lambda_dict.values()), 'bo-')
```

```
[<matplotlib.lines.Line2D at 0x7fe911769be0>]
```




```python
max(lambda_dict.items(), key=operator.itemgetter(1))[0]
```

```
6
```

Figure 1: Log-likelihood $\ell(\lambda)$ plot

From the figure above we could see that maximum value of λ is 6.

question 1.10. (10 points): draw a graphical representation of this model, which explicitly shows the random quantities and the unknown constants only.

Figure 2: Graphical representation of the model

Extra credit: mmmh ... something is amiss. the articles FZ writes have different lengths. if we model the word occurrences in each article as IID Poisson random variables with rate λ, we are implicitly assuming that the articles have the same length. why? (10 points; extra credit) and if that is true, what is the implied common length? (10 points; extra credit)

We are assuming that the FZ articles have the same length because they are from the same distribution. By using a single parameter λ we are assuming that articles have the same length. Implied common length is some constant that is the same for the all articles.

problem 2: set-up:you collect another random sample of articles penned by FZ and count how many times he uses the word however in each of the articles in your sample, $(x_1, ..., x_n)$. you also count the length of each article in your sample, $(y_1, ..., y_n)$. In this set-up, x_i is the number of times the word however appeared in the i-th article, as before, and y_i is the total number of words in the i-th article.

question 2.1. (10 points): model: let X_i denote the quantity that captures the number of times the word however appears in the i-th article. let's assume that the quantities $(X_1, ..., X_n)$ are independent and identically distributed (IID) according to a Poisson distribution with unknown parameter $\nu \cdot \frac{y_i}{1000}$,

$$p(X_i = x_i \mid y_i, \nu, 1000) = Poisson(xi \mid \nu \cdot \tfrac{y_i}{1000}) \text{ for } i = 1, ..., n.$$

using the 2-by-2 table of what's variable/constant versus what's observed/unknown, declare what's the technical nature (random variable, latent variable, known constant or unknown constant) of the quantities involved the set-up/model above: $X_1, ..., X_n$, $x_1, ..., x_n$, $y_1, ..., y_n$, ν and n.

	variable	constants
observed	$X_1, ..., X_n$	$x_1, ..., x_n$ $y_1, ..., y_n$ n
not observed		ν

question 2.2. (10 points): what is the interpretation of $\frac{y_i}{1000}$ in this model? explain.

$\frac{y_i}{1000}$ interprets the ratio between the occurrence of the number of words "however" and the length of the article. If the length of the article is longer (it has more words in the article), it is more likely to have a higher occurrence of the word "however".

question 2.3. (10 points): what is the interpretation of ν in this model? explain.

ν interprets the rate of the writing style of the author. If the writing style is good ν will be higher, otherwise, it will be lower.

question 2.4. (10 points): write the data generating process for the model above.

for $i = 1...n$ **do**
 set ν
 take y_i from dataset
 sample $x_i \sim \text{Poisson}(\nu \cdot \frac{y_i}{1000})$; assuming x_i is IID
end for

question 2.5. (10 points): define the likelihood $L(\nu) = p(\cdot \mid \cdot)$ for this model and set-up at the highest level of abstraction.

$$p(X = x \mid y_i, \nu, 1000) = \frac{e^{-\nu \cdot \frac{y_i}{1000}} \cdot (\nu \cdot \frac{y_i}{1000})^x}{x!}$$

question 2.6. (10 points): write the likelihood $L(\nu)$ for a generic sample of n articles, $(x_1, ..., x_n)$, and n articles lengths, $(y_1, ..., y_n)$.

$L(\nu) = P(X_1 = x_1 \mid y_1, \nu, 1000) \cdot P(X_2 = x_2 \mid y_2, \nu, 1000)...P(X_n = x_n \mid y_n, \nu, 1000)$

$L(\nu) = \prod_{i=1}^{n} P(X_i = x_i \mid y_i, \nu, 1000)$
$L(\nu) = \prod_{i=1}^{n} \frac{e^{-\nu \cdot \frac{y_i}{1000}} \cdot (\nu \cdot \frac{y_i}{1000})^{x_i}}{x_i!}$

question 2.7. (10 points): write the log-likelihood $\ell(\nu)$ for a generic sample of n articles, $(x_1, ..., x_n)$, and n articles lengths, $(y_1, ..., y_n)$.

$\log(L(\nu \mid x)) = \ell(\nu \mid x) = \sum_{i=1}^{n}(-\nu \cdot \frac{y_i}{1000} + x_i \log(\nu \cdot \frac{y_i}{1000}) - \log(x_i!))$

$\ell(\nu \mid x) = -\nu \cdot \sum_{i=1}^{n} \frac{y_i}{1000} + \sum_{i=1}^{n} x_i \cdot \log(\nu \cdot \frac{y_i}{1000}) - \sum_{i=1}^{n} \log(x_i!)$

question 2.8. (10 points): Simulate the number of occurrences of the word however for 5 articles using the data generating process. Assume $\nu = 10$ and coresponding article lengths $y = (1730, 947, 1830, 1210, 1100)$. Record the number of occurrences of however in each article.

As we could see from figure 4 above that the number of occurrences of the word "however" for 5 articles are: $x = (12, 10, 18, 11, 6)$

question 2.9. (10 points): write the log-likelihood $\ell(\nu)$ for the following the specific sample of occurrences you generated in the previous question and their corresponding 5 article lengths $y = (1730, 947, 1830, 1210, 1100)$.

$n = 5$
$x = (12, 10, 18, 11, 6)$
$y = (1730, 947, 1830, 1210, 1100)$

```r
require(graphics)
# Specify sample size
set.seed (47)
rnorm(5)

lambda <- list(17.3, 9.47, 18.3, 12.1, 11)

for (l in lambda) {
    cat("\nFor lambda: ", l, "\n")
    cat("x: ", rpois(1, lambda = l))
}
```

Figure 3: Code in R for question 2.8.

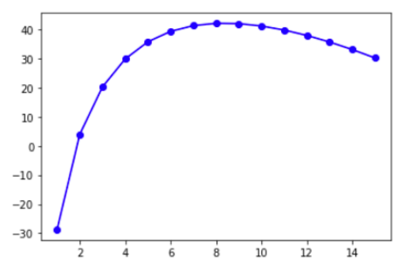
```
For lambda:  17.3
x:  12
For lambda:  9.47
x:  10
For lambda:  18.3
x:  18
For lambda:  12.1
x:  11
For lambda:  11
x:  6
```

Figure 4: Number of occurrences of "however" in each article for question 2.8.

$\ell(\nu \mid x) = -\nu \cdot \sum_{i=1}^{n} \frac{y_i}{1000} + \sum_{i=1}^{n} x_i \cdot \log(\nu \cdot \frac{y_i}{1000}) - \sum_{i=1}^{n} \log(x_i!)$
$\ell(\nu \mid x) = -\nu \cdot 6.817 + (12 \cdot log(1.73 \cdot \nu) + 10 \cdot log(0.947 \cdot \nu) + 18 \cdot log(1.83 \cdot \nu) + 11 \cdot log(1.21 \cdot \nu) + 6 \cdot log(1.1 \cdot \nu)) - 41.5$

---

question 2.10. (10 points): Plot the log-likelihood from the previous question in R. Does the maximum occur near $\nu = 10$?




```
max(ni_dict.items(), key=operator.itemgetter(1))[0]
8
```

Figure 5: Log-likelihood $\ell(\nu)$ plot

The maximum occur at $\nu = 8$.

---

question 2.11. (10 points): draw a graphical representation of this model, which explicitly shows the random quantities and the unknown constants only.

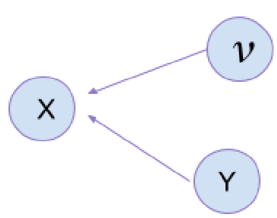


Figure 6: Graphical representation of the model

---

problem 3: set-up:you collect another random sample of articles penned by FZ and count how many times he uses the word I in each of the articles in your sample,

$(x_1, ..., x_n)$. **In this set-up,** x_i **is the number of times the word I appeared in the** i**-th article.**

---

question 3.1. (10 points): model: let X_i **denote the quantity that captures the number of times the word I appears in the** i**-th article. let** Z_i **indicate whether the** i**-th article is about politics, denoted by** $Z_i = 1$**, or not, denoted by** $Z_i = 0$**. let's assume that the quantities** $(X_1, ..., X_n)$ **are independent of one another conditionally on the corresponding values of** $Z_1, ..., Z_n$**. let's assume that the quantities** $Z_1, ..., Z_n$ **are independent and identically distributed (IID) according to a Bernoulli distribution with parameter** π**,**

$$p(Z_i \mid \pi) = Bernoulli(z_i \mid \pi) \text{ for } i = 1, ..., n.$$

let's further assume that the number of occurrences of the word I in an article about politics follows a Poisson distribution with unknown parameter $\lambda_{Politics}$**,**

$$p(X_i = x_i \mid Z_i = 1, \lambda_{Politics}) = Poisson(xi \mid \lambda_{Politics}) \text{ for } i = 1, ..., n.$$

and that the number of occurrences of the word I in an article about any other topic follows a Binomial distribution with size 1000 and unknown parameter θ_{Other}**,**

$$p(X_i = x_i \mid Z_i = 0, 1000, \theta_{Other}) = Binomial(xi \mid 1000, \theta_{Other}) \text{ for } i = 1, ..., n.$$

using the 2-by-2 table of what's variable/constant versus what's observed/unknown, declare what's the technical nature (random variable, latent variable, known constant or unknown constant) of the quantities involved the set-up/model above: $X_1, ..., X_n$**,** $x_1, ..., x_n$**,** $Z_1, ..., Z_n$**,** $z_1, ..., z_n$**,** π**,** $\lambda_{Politics}$**,** θ_{Other} **and** n**.**

---

	variable	constants
observed	$X_1, ..., X_n$	$x_1, ..., x_n$ n
not observed	$Z_1, ..., Z_n$	$z_1, ..., z_n$ $\lambda_{Politics}$ θ_{Other} π

---

question 3.2. (10 points): write the data generating process for the model above.

---

for $i = 1...n$ **do**
  set π
  set $\lambda_{Politics}$
  set θ_{Other}
  sample $z_i \sim$ Bernoulli(π); assuming z_i is IID
  if $z_i = 1$ **then**
    sample $x_i \sim$ Poisson($\lambda_{Politics}$)
  else
    sample $x_i \sim$ Binomial($1000, \theta_{Other}$)

question 3.3. (10 points): simulate 1000 values of X_i in R from the data generating process assuming $\pi = 0.3$, $\lambda_{Politics} = 30$ and $\theta_{Other} = 0.02$. Plot the values of $X_i \mid Z_i = 1$ and $X_i \mid Z_i = 0$ as two histograms on the same plot. Color the histograms by the value of Z_i so the two populations can be distinguished.

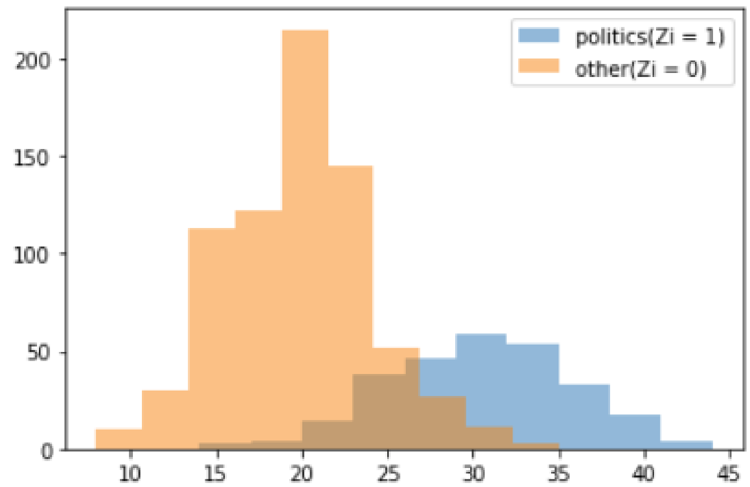


Figure 7: Histograms of two populations

question 3.4. (10 points): write the likelihood for 1 article, $L_i(\lambda_{Politics}, \theta_{Other}) = p(X_i = x_i \mid \lambda_{Politics}, \theta_{Other})$.

$$L_i(\lambda_{Politics},\ \theta_{Other}) = p(X_i = x_i \mid \lambda_{Politics}) \cdot \pi + p(X_i = x_i \mid 1000, \theta_{Other}) \cdot (1 - \pi)$$
$$L_i(\lambda_{Politics},\ \theta_{Other}) = \frac{e^{-\lambda_{Politics}} \cdot {\lambda_{Politics}}^{x_i}}{x_i!} \cdot \pi + \binom{1000}{x_i} \cdot {\theta_{Other}}^{x_i} \cdot (1 - \theta_{Other})^{1000 - x_i} \cdot (1 - \pi)$$

question 3.5. (10 points): write the likelihood $L(\lambda_{Politics}, \theta_{Other})$ for a generic sample of n articles, $(x_1, ..., x_n)$.

$$L_i(\lambda_{Politics},\ \theta_{Other}) = (p(X_1 = x_1 \mid \lambda_{Politics}) \cdot \pi + p(X_1 = x_1 \mid \theta_{Other}) \cdot (1 - \pi)) \cdot (p(X_2 = x_2 \mid \lambda_{Politics}) \cdot \pi + p(X_2 = x_2 \mid \theta_{Other}) \cdot (1 - \pi)) ... (p(X_n = x_n \mid \lambda_{Politics}) \cdot \pi + p(X_n = x_n \mid \theta_{Other}) \cdot (1 - \pi))$$
$$L_i(\lambda_{Politics},\ \theta_{Other}) = \prod_{i=1}^{n} \left(\frac{e^{-\lambda_{Politics}} \cdot {\lambda_{Politics}}^{x_i}}{x_i!} \cdot \pi + \binom{1000}{x_i} \cdot {\theta_{Other}}^{x_i} \cdot (1 - \theta_{Other})^{1000 - x_i} \cdot (1 - \pi) \right)$$

question 3.6. (10 points): write the log-likelihood $\ell(\lambda_{Politics}, \theta_{Other})$ for a generic sample of n articles, $(x_1, ..., x_n)$.

$$\ell(\lambda_{Politics}, \theta_{Other}) = \log \left(\prod_{i=1}^{n} \left(\frac{e^{-\lambda_{Politics}} \cdot {\lambda_{Politics}}^{x_i}}{x_i!} \cdot \pi + \binom{1000}{x_i} \cdot {\theta_{Other}}^{x_i} \cdot (1 - \theta_{Other})^{1000 - x_i} \cdot (1 - \pi) \right) \right)$$

question 3.7. (10 points): write the log-likelihood $\ell(\lambda_{Politics}, \theta_{Other})$ for the following specific sample of 8 articles $(12, 4, 8, 3, 3, 10, 1, 9)$.

$$\ell(\lambda_{Politics}, \theta_{Other}) = \log\Bigg(\Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{12}}{12!} \cdot \pi + \binom{1000}{12} \cdot \theta_{Other}{}^{12} \cdot (1-\theta_{Other})^{988} \cdot (1-\pi)\Bigg) \cdot \Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{4}}{4!} \cdot \pi + \binom{1000}{4} \cdot \theta_{Other}{}^{4} \cdot (1-\theta_{Other})^{996} \cdot (1-\pi)\Bigg) \cdot \Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{8}}{8!} \cdot \pi + \binom{1000}{8} \cdot \theta_{Other}{}^{8} \cdot (1-\theta_{Other})^{992} \cdot (1-\pi)\Bigg) \cdot \Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{3}}{3!} \cdot \pi + \binom{1000}{3} \cdot \theta_{Other}{}^{3} \cdot (1-\theta_{Other})^{997} \cdot (1-\pi)\Bigg) \cdot \Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{3}}{3!} \cdot \pi + \binom{1000}{3} \cdot \theta_{Other}{}^{3} \cdot (1-\theta_{Other})^{997} \cdot (1-\pi)\Bigg) \cdot \Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{10}}{10!} \cdot \pi + \binom{1000}{10} \cdot \theta_{Other}{}^{10} \cdot (1-\theta_{Other})^{990} \cdot (1-\pi)\Bigg) \cdot \Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{1}}{1!} \cdot \pi + \binom{1000}{1} \cdot \theta_{Other}{}^{1} \cdot (1-\theta_{Other})^{999} \cdot (1-\pi)\Bigg) \cdot \Bigg(\frac{e^{-\lambda_{Politics}} \cdot \lambda_{Politics}{}^{9}}{9!} \cdot \pi + \binom{1000}{9} \cdot \theta_{Other}{}^{9} \cdot (1-\theta_{Other})^{991} \cdot (1-\pi)\Bigg)\Bigg)$$

---

question 3.8. (10 points): draw a graphical representation of this model, which explicitly shows the random quantities and the unknown constants only.

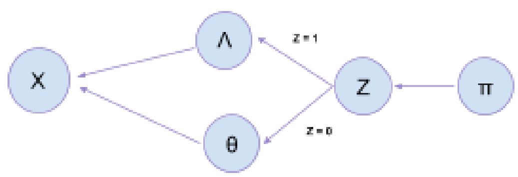


Figure 8: Graphical representation of the model

Extra credit: wait, but is it reasonable to assume that the rate λ is an unknown constant in all of our models? it seems like a stretch. (10 points; if you agree)

It is reasonable to assume that λ is an unknown constant in all of our models. Because in problem 2 the number of occurrences of the word "however" depends on the length of the article. Therefore in that case rate (λ) depends on the length. In this problem (problem 3), we didn't consider y_i parameter. We are assuming that λ is a constant as the length of the article is the same but it is not reasonable. If λ depends on y_i then it is not constant.

---

problem 4: set-up:let's go back to the simplest possible set-up for this exercise. you collect a random sample of articles penned by FZ and count how many times he uses the word and in each of the articles in your sample, $(x_1, ..., x_n)$. In this set-up, x_i is the number of times the word and appeared in the i-th article, as before.

---

question 4.1. (10 points): model: let X_i denote the quantity that captures the number of times the word and appears in the i-th article. let's assume that the quantities $(X_1, ..., X_n)$ are independent and identically distributed (IID) according to a Poisson distribution with unknown parameter Λ,

$$p(X_i = x_i \mid \Lambda = \lambda_i) = Poisson(xi \mid \lambda_i) \text{ for } i = 1, ..., n.$$

in addition, let's assume that the rate Λ is distributed according to a Gamma distribution with unknown parameters α and θ,

$$p(\Lambda = \lambda_i \mid \alpha, \theta) = Gamma(\lambda_i \mid \alpha, \theta).$$

using the 2-by-2 table of what's variable/constant versus what's observed/unknown, declare what's the technical nature (random variable, latent variable, known constant or unknown constant) of the quantities involved the set-up/model above: $X_1, ..., X_n$, $x_1, ..., x_n$, Λ, $\lambda_1, ..., \lambda_n$, α, θ and n.

---

	variable	**constants**
observed	$X_1, ..., X_n$	$x_1, ..., x_n$ n
not observed	Λ	α θ $\lambda_1, ..., \lambda_n$

---

question 4.2. (10 points): write the data generating process for the model above.

---

for $i = 1...n$ **do**
  set α
  set θ
  set $\lambda_i \sim \text{Gamma}(\alpha, \theta)$
  sample $x_i \sim \text{Poisson}(\lambda_i)$; assuming x_i is IID
end for=0

---

question 4.3. (10 points): in R simulate 1000 values from the data generating process. Assume $\alpha = 10$ and $\theta = 1$. Compute the mean and variance of the X_i.

```
N <- 1000
alpha = 10
theta = 1

# Draw N gamma distributed values
y_rgamma <- rgamma(N, shape = alpha, scale = theta)

res <- rpois(1000, lambda = y_rgamma)
mean(res)
var(res)
```

Figure 9: Simulate 1000 values from the data generating process

Mean: 10.002
Variance = 19.31731

---

question 4.4. (10 points): in R simulate 1000 values assuming $\lambda_i = 10$ for all i (ignore the Gamma distribution). Compute the mean and variance of the X_i now. How do they compare to the mean and variance you calculated in question 4.3?

```
res <- rpois(1000, lambda = 10)
mean(res)
var(res)
```

Figure 10: Simulate 1000 values $\lambda_i = 10$ (question 4.4.)

Mean: 9.971
Variance = 9.815975
Mean is similar as it is in question 4.3 while the variance is bigger in question 4.3. In question 4.3. the variance is almost two times bigger than mean. In Poisson distribution, variance is the same as mean (λ). When the variance is bigger than mean than it is called overdispersion. λ is distributed from Gamma distribution in the first case and that could have an influence on high variance.

---

question 4.5. (10 points): write the likelihood for 1 article, $L_i(\alpha,\, \theta) = p(X_i = x_i \mid \alpha, \theta)$.

$L_i(\alpha,\, \theta) = p(X_i = x_i \mid \alpha, \theta) = \frac{e^{-\lambda_i} \cdot {\lambda_i}^{x_i}}{x_i!} \cdot \frac{\theta^\alpha}{\Gamma(\alpha)} \cdot \lambda_i^{\alpha-1} \cdot e^{-\theta \cdot \lambda_i}$

---

question 4.6. (10 points): write the log-likelihood $\ell(\alpha, \theta)$ for a generic sample of n articles, $(x_1, ..., x_n)$.

$\ell(\alpha,\theta) = \log(L_i(\alpha,\, \theta)) = \log\left(\prod_{i=1}^{n} \left(\frac{e^{-\lambda_i} \cdot {\lambda_i}^{x_i}}{x_i!} \cdot \frac{\theta^\alpha}{\Gamma(\alpha)} \cdot \lambda_i^{\alpha-1} \cdot e^{-\theta \cdot \lambda_i} \right) \right)$
$\ell(\alpha,\theta) = \sum_{i=1}^{n} (-\lambda_i + x_i \log(\lambda_i) - \log(x_i!) + \alpha \log(\theta) - \log(\Gamma(\alpha)) + (\alpha - 1)\log(\lambda_i) - \theta \cdot \lambda_i)$
$\ell(\alpha,\theta) = -\sum_{i=1}^{n} \cdot \lambda_i + \sum_{i=1}^{n} x_i \log(\lambda_i) - \sum_{i=1}^{n} \log(x_i!) + n \cdot \alpha \log(\theta) - n \cdot \log(\Gamma(\alpha)) + (\alpha-1) \sum_{i=1}^{n} \log(\lambda_i) - \theta \cdot \sum_{i=1}^{n} \lambda_i$

---

question 4.7. (10 points): write the log-likelihood $\ell(\alpha, \theta)$ for the following specific sample of 8 articles $(64, 61, 89, 55, 57, 76, 47, 55)$.

$\ell(\alpha,\theta) = -\sum_{i=1}^{n} \cdot \lambda_i + \sum_{i=1}^{n} x_i \log(\lambda_i) - \sum_{i=1}^{n} \log(x_i!) + n \cdot \alpha \log(\theta) - n \cdot \log(\Gamma(\alpha)) + (\alpha-1) \sum_{i=1}^{n} \log(\lambda_i) - \theta \cdot \sum_{i=1}^{n} \lambda_i$
$\sum_{i=1}^{n} \log(x_i!) = log(64!) + log(61!) + log(89!) + log(55!) + log(57!) + log(76!) + log(47!) + log(55!) = 702.53$
$\ell(\alpha,\theta) = -\sum_{i=1}^{n} \cdot \lambda_i + 64 \cdot \log(\lambda_i) + 61 \cdot \log(\lambda_i) + 89 \cdot \log(\lambda_i) + 55 \cdot \log(\lambda_i) + 57 \cdot \log(\lambda_i) + 76 \cdot \log(\lambda_i) + 47 \cdot \log(\lambda_i) + 55 \cdot \log(\lambda_i) - 702.53 + 8 \cdot \alpha \log(\theta) - 8 \cdot \log(\Gamma(\alpha)) + (\alpha - 1) \sum_{i=1}^{n} \log(\lambda_i) - \theta \cdot \sum_{i=1}^{n} \lambda_i$

---

question 4.8. (10 points): draw a graphical representation of this model, which explicitly shows the random quantities and the unknown constants only.

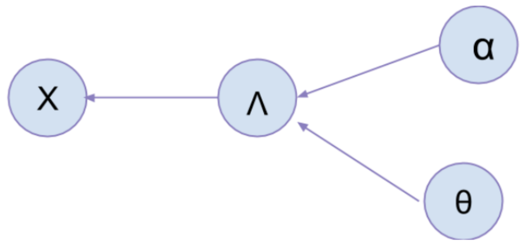


Figure 11: Graphical representation of the model

Extra credit: do you recognize the very special probability mass function you just obtained for $p(X_i = x_i \mid \alpha, \theta) = L_i(\alpha,\ \theta)$? (10 points; extra credit) excellent! you just proved a useful result: Gamma mixture of Poisson is a ... .

It is called a negative binomial distribution. It has the useful property that its variance can be greater than its mean, as the answer in question 4.3 has.

Generate samples from this distribution and verify graphically that you get the distribution looks the same as that in 4.3 (you must use appropriate parameters you identified above). (10 points; extra credit)

From the images above we could see that those two distribution are identical. Below each figure, there is mean and variance for each distribution.

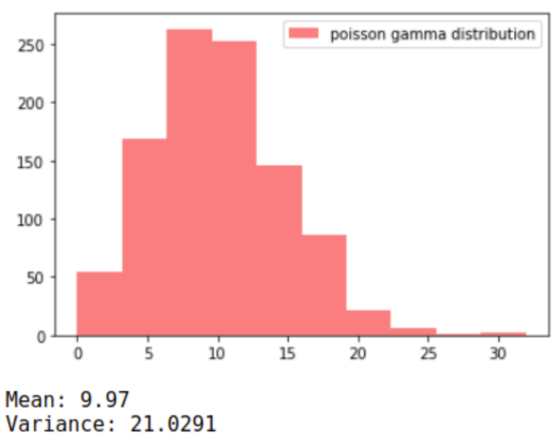


Figure 12: Distribution from question 4.3.

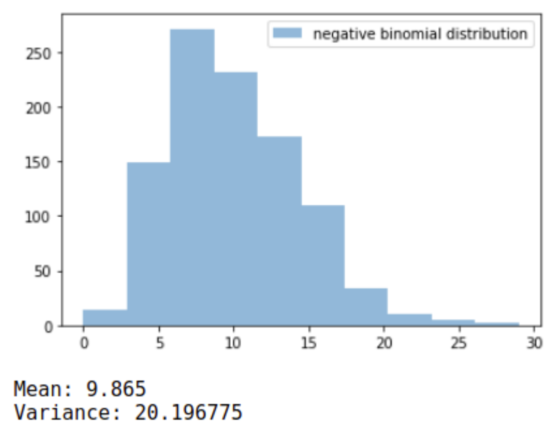


Figure 13: Negative binomial distribution