- **First Name:** Ali
- **Last Name:** Lafzi
- **Email:** alafzi@purdue.edu

Problem 1

This exercise demonstrates that probability theory is actually an extension of logic. Assume that you know that ``A implies B". That is, y information is:

$$I = \{A \implies B\}. \implies P(B|AI) = 1$$

Please answer the following questions in the space provided:

A. (4 points) $p(AB|I) = p(A|I)$.

Proof: Bayes rule: $P(AB|I) = P(A|BI)P(B|I) = \underbrace{P(B|AI)}_{1} P(A|I)$

$\implies P(AB|I) = P(A|I)$ ✓

B. If $p(A|I) = 1$, then $p(B|I) = 1$.

Proof: $\gg$ RHS $= P(B|AI)\,P(A|I) = 1 \implies P(A|BI)\,P(B|I) = 1$

$\implies$ both $P(B|I)$ & $P(A|BI)$ should be 1 ✓

C. If $p(B|I) = 0$, then $p(A|I) = 0$.

Proof: $\Rightarrow$ RHS $= P(A|BI)P(B|I) = 0 \implies P(B|AI)P(A|I) = 0$

$\xrightarrow{P(B|AI)=1} P(A|I) = 0$ ✓

D. B and C show that probability theory is consistent with Aristotelian logic. Now, you will discover how it extends it. Show that if B is tru
A becomes more plausible, i.e.

$$p(A|BI) \geq p(A|I).$$

Proof: B is true $\implies P(B|I) \geq 0$

Bayes rule: $P(A|BI) = \dfrac{P(AB|I)}{P(B|I)} \overset{\text{from parts (a)}}{=\!=} \dfrac{P(A|I)}{P(B|I)}$

$\implies P(A|I) = P(A|BI)\underbrace{P(B|I)}_{\leq 1} \implies P(A|BI) \geq P(A|I)$ ✓

E. Give at least two examples of D that apply to various scientific fields. To get you started, here are two examples:

- A: It is raining. B: There are clouds in the sky. Clearly, $A \implies B$. D tells us that if there are clouds in the sky, raining becomes r
plausible.
- A: General relativity. B: Light is deflected in the presence of massive bodies. Here $A \implies B$. Observing that B is true makes A
plausible.

Answer:

- A: The fluid has a high viscosity.

  B: The wall friction in the channel is high.

  $\Rightarrow$ If the wall friction in the channel is high, it is more likely that the fluid has a high viscosity.

- A: We have a large dataset.

  B: We are less likely to over fit the data.

  $\Rightarrow$ If we are less likely to overfit the data, it is more plausible that we have a large dataset.

F. Show that if A is false, then B becomes less plausible, i.e.:

not A is true $\Leftrightarrow 0 \le P(\neg A|I) \le 1$

$$p(B|\neg AI) \le p(B|I).$$

Proof:

Bayes rule: $P(B|\neg AI) = \dfrac{P(B, \neg A|I)}{P(\neg A|I)} = \dfrac{P(\neg A|BI)\,P(B|I)}{P(\neg A|I)} = \dfrac{[1-P(A|BI)]\,P(B|I)}{P(\neg A|I)}$

$= \dfrac{P(B|I) - P(AB|I)}{P(\neg A|I)} \underset{\text{from part (a)}}{=} \dfrac{P(B|I) - P(A|I)}{P(\neg A|I)} = \dfrac{P(B|I) - 1 + P(\neg A|I)}{P(\neg A|I)} = P(B|\neg AI)$

$\Rightarrow 1 - P(B|I) = \underbrace{P(\neg A|I)}_{\le 1}\,[1 - P(B|\neg AI)] \Rightarrow 1 - P(B|I) \le 1 - P(B|\neg AI) \Rightarrow P(B|I) \ge P(B|\neg AI)$

✓

G. Can you think of an example of scientific reasoning that involves F? For example: A: It is raining. B: There are clouds in the sky. F te that if it is not raining, then it is less plausible that there are clouds in the sky.

Answer: A. He is injured.

B. He has a low chance of winning the competition.

If he is not injured, it is less possible that he has a low chance of winning the competition.

"All swans are white"

H. Do D and F contradict Karl Popper's principle of falsification, "A theory in the empirical sciences can never be proven, but it can be fa
meaning that it can and should be scrutinized by decisive experiments." → it is possible to observe a black swan

Answer: let us refer to the famous example of falsibility as follows:

A implies B, A = not all swans are white, B = we can find black swans

According to part D, if $\underbrace{\text{we can find black swans}}_{\text{B is true}}$, it is more likely that $\underbrace{\text{not all swans are white}}_{\text{A is more plausible}}$. This is in agreement with the principle of falsification.

F states that if $\underbrace{\text{all swans are white}}_{\text{A is false}}$, there is $\underbrace{\text{less chance that we can find black swans}}_{\text{B is less likely}}$.

This also does not contradict the principle.

Problem 2

Disclaimer: This example is a modified version of the one found in a 2013 lecture on Bayesian Scientific Computing taught by Prof. Nicl
Zabaras. I am not sure where the original problem is coming from.

We are tasked with assessing the usefulness of a tuberculosis test. The prior information I is:

> The percentage of the population infected by tuberculosis is 0.4%. We have run several experiments and determined that:
>
> - If a tested patient has the disease, then 80% of the time the test comes out positive.
> - If a tested patient does not have the disease, then 90% of the time the test comes out negative.

To facilitate your analysis, consider the following logical sentences concerning a patient:

> A: The patient is tested and the test is positive.

> B: The patient has tuberculosis.

A. Find the probability that the patient has tuberculosis (before looking at the result of the test), i.e., $p(B|I)$. This is known as the base the prior probability.

Answer: $= P(B|I) = 0.4\% = 0.004$

B. Find the probability that the test is positive given that the patient has tuberculosis, i.e., $p(A|B, I)$.

Answer: $= P(A|BI) = 80\% = 0.8$

C. Find the probability that the test is positive given that the patient does not have tuberculosis, i.e., $p(A|\neg B, I)$.

Answer: $= P(A|\neg B I) = 1 - P(\neg A|\neg B I) = 1 - 0.9 = 0.1$

D. Find the probability that a patient that tested positive has tuberculosis, i.e., $p(B|A, I)$.

Answer: $P(B|AI) = \frac{P(A|BI)\,P(B|I)}{P(A|I)}$ $P(A|I) = P(A,B|I) + P(A,\sim B|I)$

$= P(A|BI)\,P(B|I) + P(A|\sim BI)\,[1 - P(B|I)] = 0.8 \times 0.004 + 0.1 \times (1 - 0.004) = 0.1028$

$\Rightarrow P(B|AI) = \frac{0.8 \times 0.004}{0.1028} = 0.03112$

E. Find the probability that a patient that tested negative has tuberculosis, i.e., $p(B|\neg A, I)$. Does the test change our prior state of knc about about the patient? Is the test useful?

Answer:

$P(B|I) = P(B,A|I) + P(B,\sim A|I) = P(B|AI)\,P(A|I) + P(B|\sim AI)\,P(\sim A|I) \Rightarrow 0.004 = 0.03112 \times 0.1028 + P(B|\sim AI) \times (1 - 0.1028)$

$\Rightarrow P(B|\sim AI) = 0.000893$ $\Rightarrow P(B|\sim AI) < P(B|I)$ $\longrightarrow$ therefore, our prior state of knowledge is updated and the test is useful.

F. What would a good test look like? Find values for

$$p(A|B, I) = p(\text{test is positive}|\text{has tuberculosis}, I),$$

and

$$p(A|\neg B, I) = p(\text{test is positive}|\text{does not have tuberculosis}, I),$$

so that

$$p(B|A,I) = p(\text{has tuberculosis}|\text{test is positive}, I) = 0.99.$$

There are more than one solutions. How would you pick a good one? Thinking in this way can help you set goals if you work in R&D. If y time, try to figure out whether or not there exists such an accurate test for tuberculosis

Answer: $P(B|AI) = \dfrac{P(AB|I)}{P(A|I)} = \dfrac{P(A|BI)\,P(B|I)}{P(AB|I) + P(A,\sim B|I)} = \dfrac{P(A|BI)\,P(B|I)}{P(A|BI)\,P(B|I) + P(A|\sim BI)\,P(\sim B|I)}$

$\Rightarrow 0.99 = \dfrac{0.004 \times P(A|BI)}{0.004 \times P(A|BI) + 0.996 \times P(A|\sim BI)} \Rightarrow P(A|BI) = 24651\, P(A|\sim BI)$

A good test is one having very high $P(A|BI)$ and very low $P(A|\sim BI)$ so that the test is accurate ($p(B|AI)$ is high). This can be done by either increasing $P(A|BI)$ or decreasing $P(A|\sim BI)$. For instance, if $P(A|BI)=0.95$ $\xrightarrow{P(B|AI)=0.99}$ $P(A|\sim BI) = 0.0000385$

Problem 3

Let A and B be independent conditional on I. Prove that:

$$A \perp B|I \iff p(AB|I) = p(A|I)p(B|I).$$

Hint: Use the fact that $A \perp B|I$ means that $p(A|B,I) = p(A|I)$ and $p(B|A,I) = p(B|I)$.

Answer: ① if $A \perp B|I \Rightarrow P(AB|I) = P(A|I)P(B|I)$

$P(AB|I) = P(A|BI)\,P(B|I) \xrightarrow{\text{based on definision}} P(AB|I) = P(A|I)\,P(B|I)$

Similarly: $P(BA|I) = P(B|AI)\,P(A|I) = P(B|I)\,P(A|I)$

② if $P(AB|I) = P(A|I)\,P(B|I) \Rightarrow A \perp B|I$

$P(AB|I) = P(A|I)\,\cancel{P(B|I)} \overset{\text{Bayes rule}}{=\!=} P(A|BI)\,\cancel{P(B|I)} \Rightarrow P(A|I) = P(A|BI) \Rightarrow A \perp B|I$

similarly: $P(BA|I) = \cancel{P(A|I)}\,P(B|I) = P(B|AI)\,\cancel{P(A|I)} \Rightarrow P(B|I) = P(B|AI) \Rightarrow B \perp A|I$

Problem 4

Let X be a continuous random variable and $F(x) = p(X \leq x)$ be it's cumulative distribution function. Using only the basic rules of probability, prove that:

A. The CDF starts at 0 and goes up to 1:

$$F(-\infty) = 0 \text{ and } F(\infty) = 1.$$

Proof:

$F(-\infty) = p(X \leq x = -\infty)$ → this is an impossible event since the value of random variable cannot be less than $-\infty$. $\overset{i.e.}{\Rightarrow} F(-\infty) = 0$

$F(+\infty) = p(X \leq x = +\infty)$ → $+\infty$ is the maximum value for the random variable. Therefore, this event contains all the possible values for x and is a certain event. $\overset{i.e.}{\Rightarrow} F(\infty) = 1$

B. $F(x)$ is a monotonically increasing function of x, i.e.,

$$x_1 \leq x_2 \implies F(x_1) \leq F(x_2).$$

Proof:

$F(x_2) = P(X \leq x_2) = \underbrace{P(X \leq x_1)}_{F(x_1)} + \underbrace{P(x_1 \leq X \leq x_2)}_{\geq 0} - \underbrace{P(X \leq x_1 \text{ \& } x_1 \leq X \leq x_2)}_{= 0 \text{ impossible event since } x_1 \leq x_2}$

$\Rightarrow F(x_2) \geq F(x_1)$

C. The probability of X being in the interval $[x_1, x_2]$ is:

$$p(x_1 \leq X \leq x_2 | I) = F(x_2) - F(x_1).$$

Proof:

As shown above: $P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$

Problem 5

Let X be a random variable. Prove that:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Proof:

by definition: $V[X] = E[(x - E[x])^2] = E(x^2 + E[x]^2 - 2x\,E[x]) \overset{\text{from properties}}{=\!=} E[x^2] + E[E[x]^2] - 2E[x\,E[x]]$

Expectation of a random variable is a constant number and is deterministic and not random anymore. Hence, it can be pulled out of the expectation expression. For the same reason, expectation of expectation of a random variable is the corresponding expectation itself.

$\Rightarrow V(x) = E(x^2) + E[x]^2 - 2E[x]E(x) = E(x^2) - E[x]^2$ ✓

Problem 6

Hint: Before attempting this example, make sure you understand the Lecture 5 examples. You basically have to repeat the same procec

The San Andreas fault extends through California forming the boundary between the Pacific and the North American tectonic plates. It caused some of the major earthquakes on Earth. We are going to focus on Southern California and we would like to assess the probab major earthquake, defined as an earthquake of magnitude 6.5 or greater, during the next ten years.

A. The first thing we are going to do is go over a database of past earthquakes that have occured in Southern California and collect the data. We are going to start at 1900 because data before that time may are unreliable. Go over each decade and count the occurence of earthquake (i.e., count the number of organge and red colors in each decade). We have done this for you.

```
eq_data = np.array([
    0, # 1900-1909
    1, # 1910-1919
    2, # 1920-1929
    0, # 1930-1939
    3, # 1940-1949
    2, # 1950-1959
    1, # 1960-1969
    2, # 1970-1979
    1, # 1980-1989
    4, # 1990-1999
    0, # 2000-2009
    2 # 2010-2019
])
fig, ax = plt.subplots()
ax.bar(np.linspace(1900, 2019, eq_data.shape[0]), eq_data, width=10)
ax.set_xlabel('Decade')
ax.set_ylabel('# of major earthquakes in Southern CA');
```

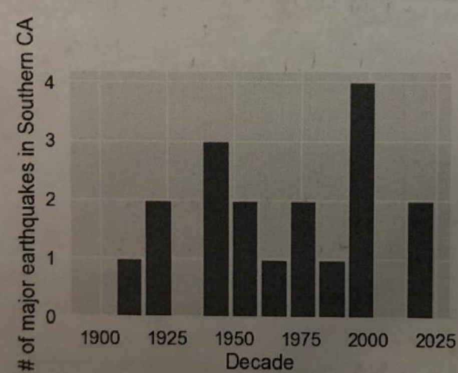
$P(x=0) = \frac{3}{12}$

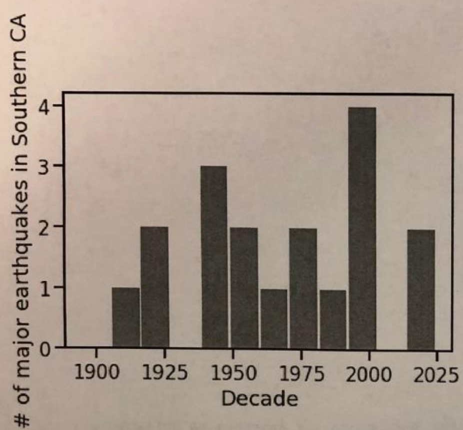
$P(x=1) = \frac{3}{12}$

$P(x=2) = \frac{4}{12}$ $\Rightarrow$ they add up to 1 ✓

$P(x=3) = \frac{1}{12}$

$P(x=4) = \frac{1}{12}$





B. The right way to model the number of earthquakes X_n in a decade n is using a Poisson distribution with unknown rate parameter λ

$$X_n|\lambda \sim \text{Poisson}(\lambda).$$

Here we have $N = 12$ observations, say $x_{1:N} = (x_1, \ldots, x_N)$ (stored in `eq_data` above). Find the *joint probability* (otherwise knowr likelihood) $p(x_{1:N}|\lambda)$ of these random variables.

Answer: $P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$ poisson distribution Since $x_{1:N}$ are independent random variables, we can write: $p(x_{1:N}|\lambda) = \prod_{n=1}^{N=12} p(x_n|\lambda) = \prod_{n=1}^{N=12} \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$

$$= \frac{\lambda^{\sum_{n=1}^{N=12} x_n} e^{-N\lambda}}{\prod_{n=1}^{N=12} x_n!}$$

C. The rate parameter λ (number of major earthquakes in per ten years) is positive. What prior distribution should we assign to it. A sui choice here is to pick a Gamma, see also the scipy.stats page for the Gamma. We write:

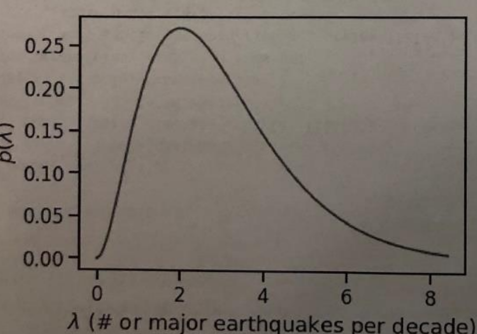
$$\lambda \sim \text{Gamma}(\alpha, \beta),$$

where α and β are positive *hyper-parameters* that we have to set to represent our prior state of knowledge. The PDF is:

$$p(\lambda) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)},$$

where we are not conditioning on α and β because they should be fixed numbers. Use the code below to pick some some reasonable \ for α and β.

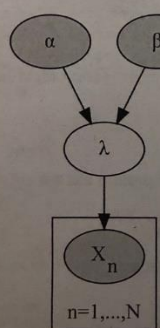
```
import scipy.stats as st
alpha = 3.0 # Pick them here
beta = 1.0  # Pick them here
lambda_prior = st.gamma(alpha, scale=1.0 / beta) # Make sure you understand why scale = 1 / beta
lambdas = np.linspace(0, lambda_prior.ppf(0.99), 100)
fig, ax = plt.subplots()
ax.plot(lambdas, lambda_prior.pdf(lambdas))
ax.set_xlabel('$\lambda$ (# or major earthquakes per decade)')
ax.set_ylabel('$p(\lambda)$');
```



α and β are chosen such that we see high probability around $x=2$ (where the peak is slightly shifted to the left) to match better with what we observed in part A.

D. Use the package `graphviz` to draw a graphical model representing the situation. Make sure you use the plate notation. Below, we ha introduced all the nodes you are going to need, but some nodes should be "observed" (observed nodes are filled), a lot of edges are mis and you need to use the plate notation (see Lecture 5).

```
from graphviz import Digraph
import os
gc = Digraph('hw01_p6_1')
gc.node('alpha', label='<&alpha;>', style='filled')
gc.node('beta', label='<&beta;>', style='filled')
gc.node('lambda', label='<&lambda;>')
#gc.node('Xn', label='<X<sub>n</sub>>')
with gc.subgraph(name='cluster_0') as sg:
    sg.node('Xn', label='<X<sub>n</sub>>', style='filled')
    sg.attr(label='n=1,...,N')
    sg.attr(labelloc='b')
gc.edge('alpha','lambda')
gc.edge('beta','lambda')
gc.edge('lambda', 'Xn')
gc
```



E. Show that the posterior of λ conditioned on $x_{1:N}$ is also a Gamma, but with updated hyperparameters. Hint: When you write down tl posterior of λ you can drop any multiplicative term that does not depend on it as it will be absorbed in the normalization constnat. This simplify the notation a little bit.

Answer: $p(\lambda|x_{1:N}) = \dfrac{p(x_{1:N}|\lambda)\,p(\lambda)}{\underbrace{p(x_{1:N})}_{cte}}$ $\quad p(x_{1:N}|\lambda) = \dfrac{\lambda^{\sum_{n=1}^{N=12} x_n} e^{-N\lambda}}{\prod_{n=1}^{N=12} x_n!}$ $\quad P(\lambda) = \dfrac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}$

$$\Rightarrow P(x_{1:N}|\lambda)P(\lambda) = \frac{\beta^{\alpha} e^{-\lambda(N+\beta)} \lambda^{\sum_{n=1}^{N=12} x_n + \alpha - 1}}{\prod_{n=1}^{N=12} x_n!\,\Gamma(\alpha)} = \frac{\lambda^{\sum_{n=1}^{N=12} x_n + \alpha - 1} e^{-\lambda(N+\beta)} \beta^{\alpha + \sum_{n=1}^{N=12} x_n}}{\prod_{n=1}^{N=12} x_n!\,\Gamma(\alpha)\,\beta^{\sum_{n=1}^{N=12} x_n}}$$

$$\alpha' = \sum_{n=1}^{N=12} x_n + \alpha \;\&\; \beta' = N+\beta \qquad \frac{\text{Gamma}(\alpha', \beta')}{\prod_{n=1}^{N=12} x_n!\,\Gamma(\alpha)\,\beta^{\sum_{n=1}^{N=12} x_n}}$$

→ so this posterior is also a gamma distribution with a normalization constant.

F. Prior-likelihood pairs that result in a posterior with the same form as the prior as known as conjugate distributions. Conjugate distribı are your only hope for analytical Bayesian inference. As a sanity check, look at the wikipedia page for conjugate priors, locate the Poiss Gamma pair and verify your answer above. This is verified. wikipedia also says: α and β are the prior hyperparameters and $\alpha + \sum_{i=1}^{n} x_i$ & $\beta + n$ are the posterior hyperparameters.

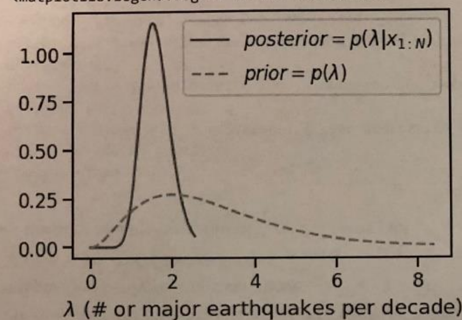
G. Plot the prior and the posterior of λ on the same plot.

```
alpha_post = alpha
for i in range(len(eq_data)):
  alpha_post = alpha_post + eq_data[i] # Your expression for alpha posterior here
#print (alpha_post)
beta_post = beta + len(eq_data) # Your expression for beta posterior here
#print (beta_post)
lambda_post = st.gamma(alpha_post, scale=1.0 / beta_post)
lambdas = np.linspace(0, lambda_post.ppf(0.99), 100)
fig, ax = plt.subplots()
ax.plot(lambdas, lambda_post.pdf(lambdas), label='$posterior=p(\lambda|x_{1:N})$')
ax.set_xlabel('$\lambda$ (# or major earthquakes per decade)')
#ax.set_ylabel('$p(\lambda|x_{1:N})$');

lambdas = np.linspace(0, lambda_prior.ppf(0.99), 100)
```

```
ax.plot(lambdas, lambda_prior.pdf(lambdas), label='$prior=p(\lambda)$', linestyle='--')
ax.set_xlabel('$\lambda$ (# or major earthquakes per decade)')
#ax.set_ylabel('$prior=p(\lambda)$');
ax.legend()
```

```
<matplotlib.legend.Legend at 0x7f8ede7b8438>
```



I. Let's work out the predictive distribution for the number of major earthquakes during the next decade. This is something that we did r class, but it will appear again and again in future lectures. Let X be the random variable corresponding to the number of major eathquа during the next decade. We need to calculate:

$$p(x|x_{1:N}) = \text{our state of knowledge about } X \text{ after seeing the data.}$$

How do we do this? We just use the sum rule:

$$p(x|x_{1:N}) = \int_0^\infty p(x|\lambda, x_{1:N})p(\lambda|x_{1:N})d\lambda = \int_0^\infty p(x|\lambda)p(\lambda|x_{1:N})d\lambda,$$

where going from the middle step to the rightmost one we used the assumption that the number of earthquakes occuring in each deca independent. Carry out this integral and show that it will give you the negative Binomial distribution $\mathrm{NB}(r, \theta)$, see also the scipy.stats ı with parameters

$$r = \alpha + \sum_{n=1}^N x_n, \qquad r = \alpha \qquad \theta = \frac{1}{\beta + 1}$$

and

$$\theta = \frac{1}{\beta + N}.$$

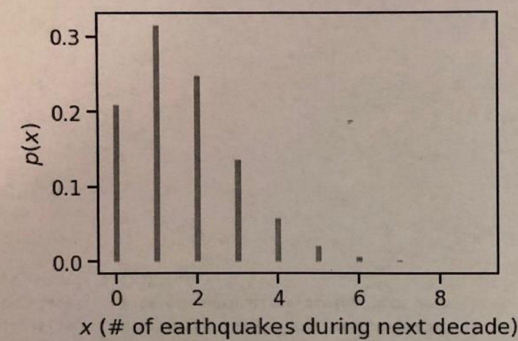
Answer: $p(x|x_{1:N}) = \int_0^\infty \underbrace{\frac{\beta'^{-\alpha'} e^{-\beta'\lambda} \lambda^{\alpha'-1}}{\Gamma(\alpha')}}_{Gamma(\alpha';\beta')} \underbrace{\frac{\lambda^{x} e^{-\lambda}}{x!}}_{poisson(\lambda)} d\lambda = \frac{\beta'^{-\alpha'}}{x!\,\Gamma(\alpha')} \int_0^\infty e^{-\lambda(\beta'+1)} \lambda^{\alpha'+x-1} d\lambda$

based on definition $\frac{\left(\frac{1-\theta}{\theta}\right)^r}{x!\,\Gamma(r)} \int_0^\infty e^{\frac{-\lambda}{\theta}} \lambda^{r+x-1} d\lambda = \frac{(1-\theta)^r}{\theta^r x!\,\Gamma(r)} \theta^{r+k}\,\Gamma(r+k)$ → This is negative binomial distribution according to Wikipedia.

J. Plot the predictive distribution $p(x|x_{1:N})$.

```
r = alpha_post # Your expression for r here
theta = 1./(beta_post + 1) # Your expression for theta here
X = st.nbinom(r, 1.0 - theta) # Please pay attention to the fact that the wiki and scipy.stats
                              # use slightly different definitions
fig, ax = plt.subplots()
xs = range(0, 10)
ax.vlines(xs, 0, X.pmf(xs), colors='b', lw=5, alpha=0.5)
ax.set_xlabel('$x$ (# of earthquakes during next decade)')
ax.set_ylabel('$p(x)$');
```

⇨



K. What is the probability that at least one major earthquake will occur during the next decade?

Answer: $p(x \geq 1) = 1 - p(x \leq 0) = 1 - F(x=0)$

$\rightarrow CDF(0)$

```
print (1-X.cdf(0))
```

```
0.7890781124491307
```

L. What is the probability that at least one major earthquake will occur during the next two decades?

Answer:

This event is equivalent to the event that "no earthquake occurs in the next decade and the following decade", is not true

In other words, it is: $1 - p(X=0, Y=0) \overset{\text{since } X \perp Y}{=\!=} 1 - p(X=0)\,p(Y=0)$

$= 1 - F(X=0)\,F(Y=0) = 1 - [CDF(0)]^2$

```
#xs = range(0, 20)
print (1-X.cdf(0)**2)
```

```
0.9555119573519785
```

M. Find a 95% prediction interval for λ.

```
# Write your code here and print() your answer
lambda_low = X.ppf(0.025)
lambda_up = X.ppf(0.975)
print('Number of earthquakes is in [{0:1.2f}, {1:1.2f}] with 95% probability'.format(theta_low, theta_up))
```

```
Number of earthquakes is in [0.00, 5.00] with 95% probability
```

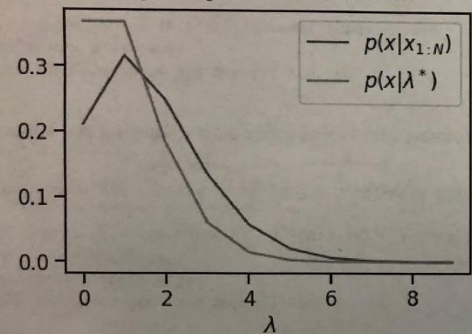
N. Find the λ that minimizes the absolute loss (see lecture), call it λ^*_N. Then, plot the fully Bayesian predictive $p(x|x_{1:N})$ to $p(x|\lambda^*_N)$.

```
# Write your code here and print() your answer
lambda_star = X.median()
print('Lambda_star = {0:1.2f}'.format(lambda_star))

lambdas = range(0, 10)
fig, ax = plt.subplots()
#ax.plot([theta_true], [0.0], 'o', markeredgewidth=2, markersize=10, label='True value')
ax.plot(lambdas, X.pmf(lambdas), label=r'$p(x|x_{1:N})$')
#ax.plot(theta_star_01, 0, 'x', markeredgewidth=2, label=r'$\theta^*_{01}$')
#ax.plot(theta_star_2, 0, 's', markeredgewidth=2, label=r'$\theta^*_{2}$')
#ax.plot(lambda_star, 0, 'd', markeredgewidth=2, label=r'$\lambda^*$')
```

```
ax.plot(lambdas, st.poisson(lambda_star).pmf(lambdas), label=r'$p(x|\lambda^*)$')
ax.set_xlabel(r'$\lambda$')
#ax.set_title('$N={0:d}$'.format(N))
plt.legend(loc='best')
```

```
Lambda_star = 1.00
<matplotlib.legend.Legend at 0x7f8edc2b25c0>
```



-End-