Modeling count data

Stat 340: Bayesian Statistics

1. Inference for count data

(Problem topics 1-4)

Example

I get A LOT of potential scam calls

Let's model the number of scam calls I receive in a day

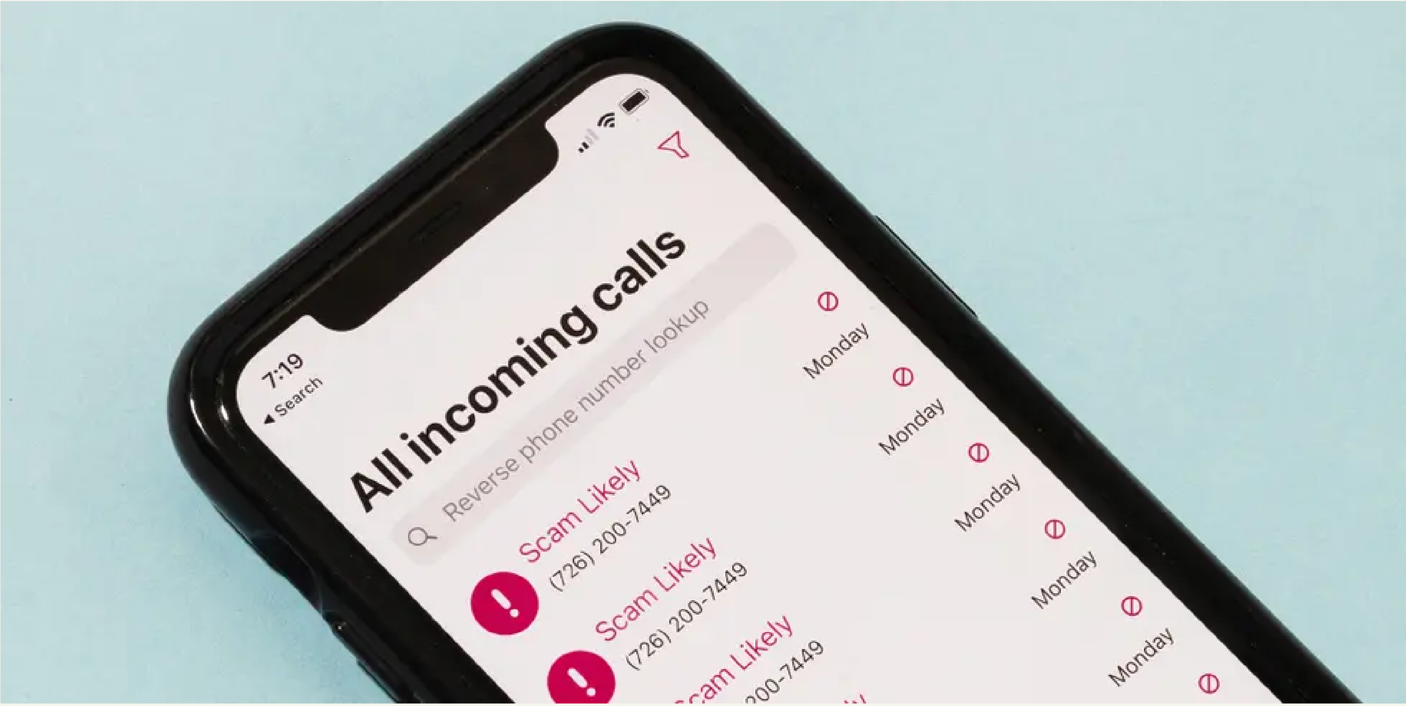


Image credit: Wirecutter

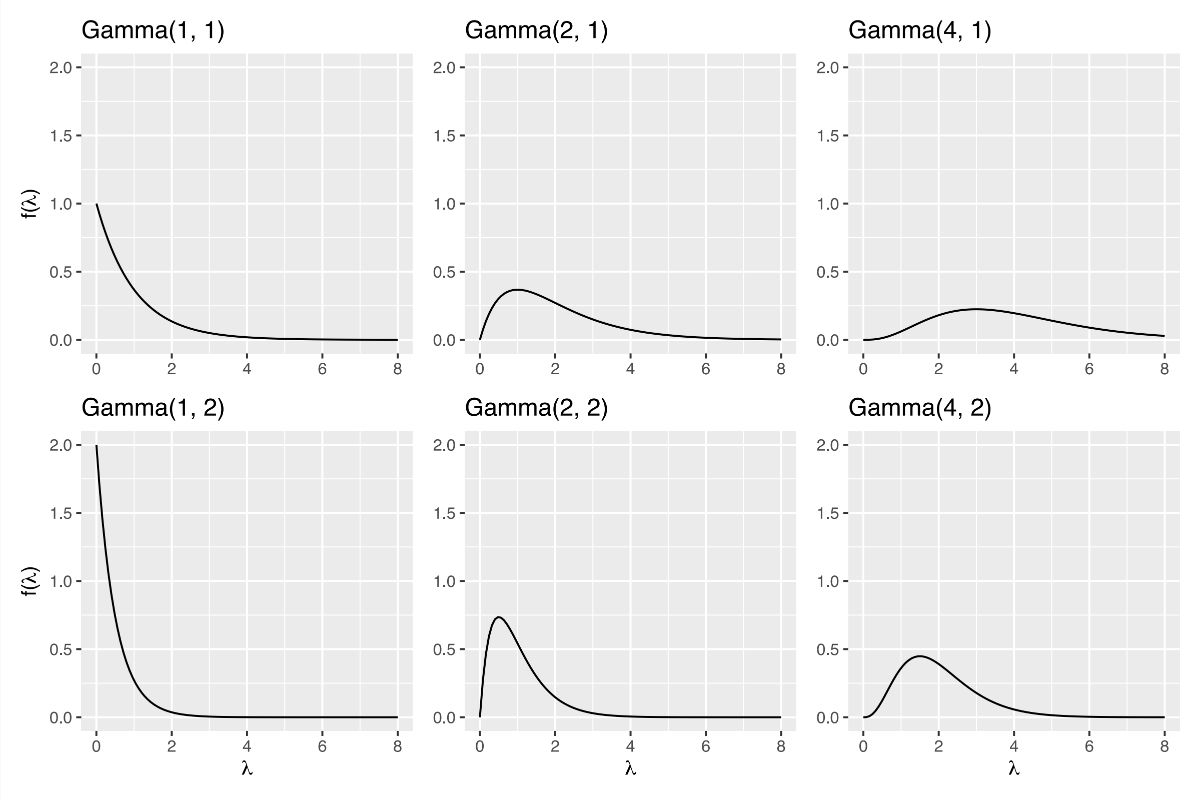
Modeling choices

To model the number of scam calls, we need...

1. A likelihood
2. A prior
3. Data

Gamma distributions

- $a = 1 \Rightarrow$ exponential
- $a > b \Rightarrow$, mean > 1
- $a < b \Rightarrow$, mean < 1
- as a increases relative to b, skew $\downarrow$ and variance $\uparrow$



`bayesrules::plot_gamma()`

Your turn 1

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with PMF

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \ x = 0, 1, 2, \ldots$$

1. Write down the likelihood function, $f(x_1, \ldots, x_n | \lambda)$.

2. Suppose that you decide to use a Gamma(a,b) prior distribution for λ with PDF $\pi(\lambda) = \dfrac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}, \ \lambda > 0.$

   Find the posterior density of λ.

3. Is the gamma prior a conjugate family to the Poisson likelihood?

04:00

Tuning a gamma prior

Set up a system of equations to solve with any two characteristics (not just quantiles)

Mean $$E(\lambda) = \frac{a}{b}$$

Mode $$\text{mode}(\lambda) = \frac{a-1}{b} \text{ if } a, \geq 1; 0 \text{ if } 0 \leq a < 1$$

Variance $$\text{Var}(\lambda) = \frac{a}{b^2}$$

Parameter solver

If you want to use two quantiles, then you can use `gamma_solver()`

```
devtools::source_gist("https://gist.github.com/aloy/9d0385363663a28bd8eccf85bf
```

```
gamma_solver(.values = c(10, 20), .probs = c(.05, .95), .guess = c(5, 1))
##         a         b
## 22.941774  1.567326
```

`.values` - values of two quantiles `.probs` - what quantiles did you specify?
`.guess` - initial guess for a and b

Your turn 2

My best guess is that this rate...

- is expected to be around 5 calls per day
- could reasonably range from 2 to 7 calls per day

What gamma prior would you use?

03:00

Updating λ

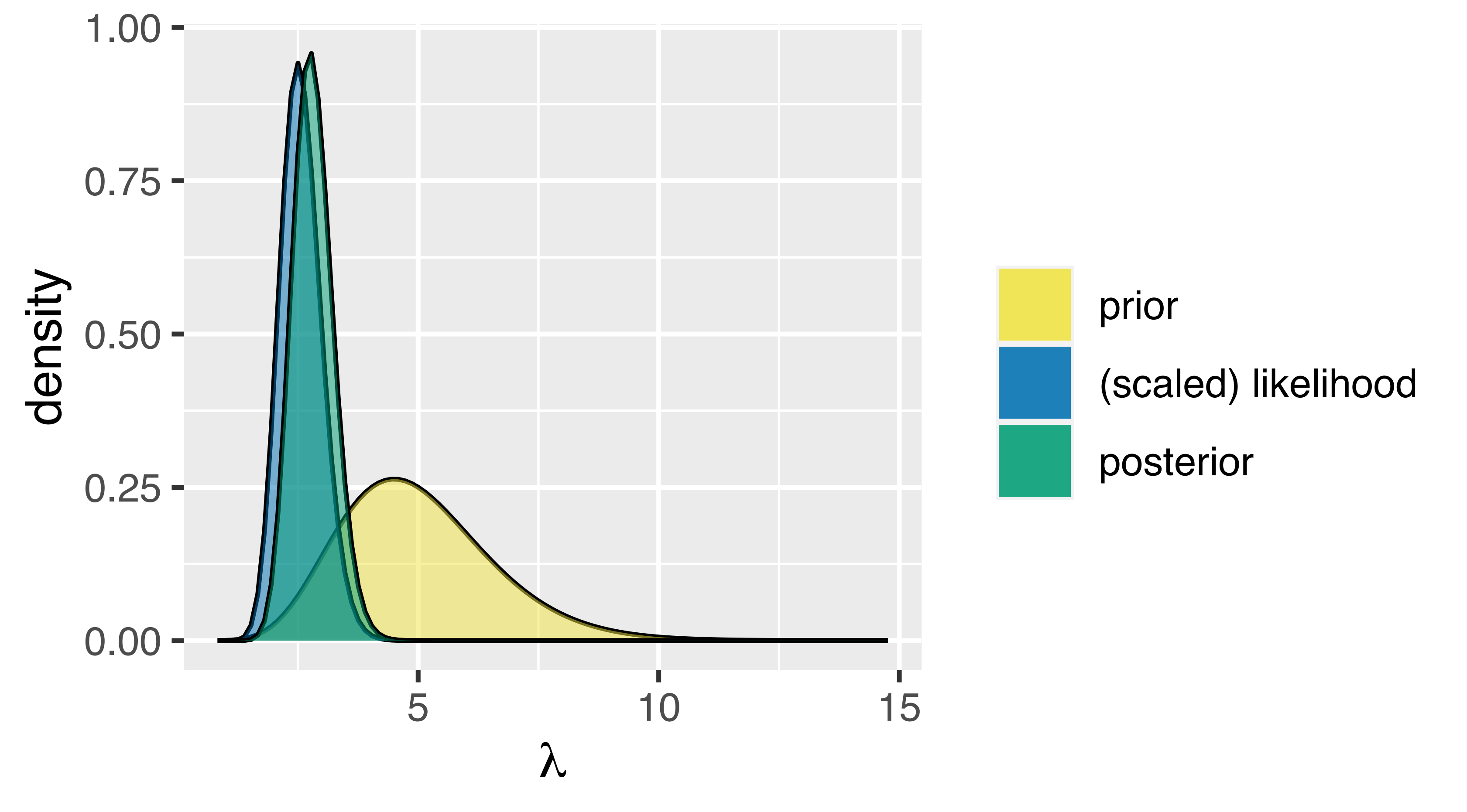
Over the first two weeks of the term, I recorded the number of scam calls I received:

```
2 1 5 5 1 3 3 1 3 3 0 6 1 1
```

Summary stats: $$\sum_{i=1}^{14} y_i = 35 \qquad n = 14$$

So our posterior distribution for λ is...

Updating λ



Your turn 3

1. How many scam calls should I expect on average in a typical day?
2. How many scam calls should I expect tomorrow?
3. How many scam calls should I expect in the next week (7 days)?