

# Gamma-Poisson Model and Posterior Inference.

Elena Moltchanova

STAT314/461-2021S1

## Sandwiches:

Table 1: Sandwich Demand

Day	x
Monday	50
Tuesday	65
Wednesday	72
Thursday	63
Friday	70

How to estimate the average daily demand for sandwiches and to provide some uncertainty bounds for them?

## Poisson Likelihood

Let  $x_i$  denote the number of sandwiches sold on day  $i$ , then

$$x_i | \mu \sim \text{Pois}(\mu), \quad \text{for } i = 1, \dots, n.$$

where  $n$  is the total number of recorded days, and  $\mu$  is the Poisson intensity parameter, corresponding to the expected number of sandwiches sold on any given day.

## Classical Analysis: Likelihood

$$f(x_i|\mu) = \frac{\mu^{x_i} e^{-\mu}}{x_i!}.$$

## Classical Analysis: Joint Likelihood

$$f(\mathbf{x}|\mu) = f(x_1, \dots, x_n|\mu) = \prod_i f(x_i|\mu) = \prod_i \frac{\mu^{x_i} e^{-\mu}}{x_i!} = \frac{\mu^{\sum_i x_i} e^{-n\mu}}{\prod_i x_i!}.$$

## Maximizing the likelihood:

- ▶ Log the likelihood
- ▶ Find the first derivative w.r.t.  $\mu$  and solve for 0. Call the solution  $\hat{\mu}$
- ▶ Find the second derivative and check that it is negative (for maximum) at  $\hat{\mu}$ .
- ▶ The solution  $\hat{\mu}$  is the Maximum Likelihood Estimate.

Maximizing the likelihood:

## Poisson MLE:

$$\hat{\mu} = \bar{x} = 64.$$



## Bayesian Derivation: What is the prior?

The Poisson intensity parameter  $\mu$  is positive.

$$\mu|a, b \sim \text{Gamma}(a, b),$$

i.e.

$$f(\mu|a, b) = \frac{b^a}{\Gamma(a)} \mu^{a-1} e^{-b\mu} \quad \text{for } \mu > 0,$$

where  $\Gamma(a)$  is a gamma function.

# Applying Bayes' Theorem

## Posterior distribution

$$\mu | \mathbf{x}, a, b \sim \text{Gamma}(a + \sum_i x_i, b + n).$$

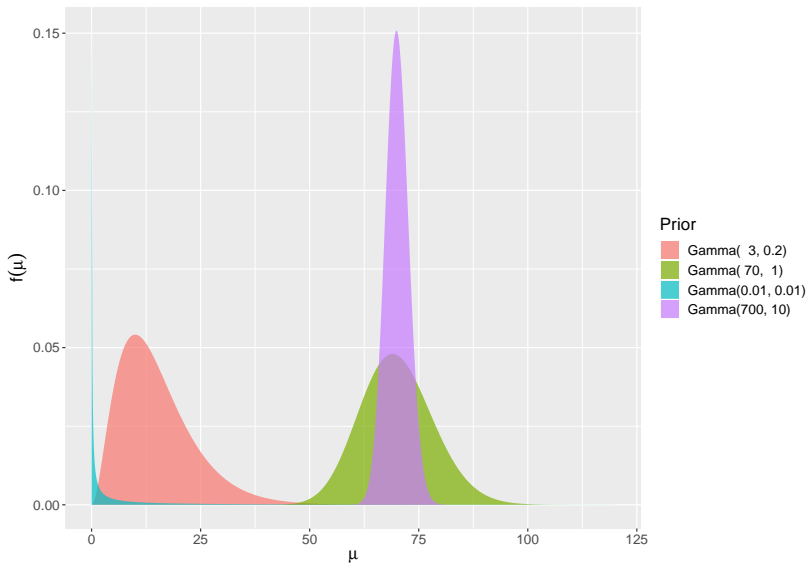
## Conjugate posterior distributions:

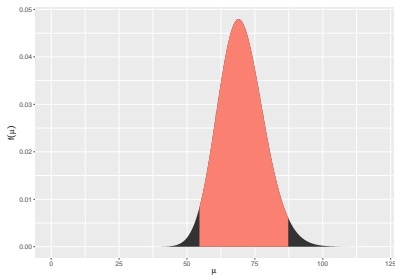
In the previous derivation, we combined a **gamma** prior with a Poisson likelihood to obtain a **gamma** posterior distribution.

Therefore, gamma distribution is a **conjugate** prior for the Poisson likelihood.

Conjugate priors are often selected for convenience, although with the advance of numerical methods and computer technology they have become less important.

## Gamma distributions: $\text{Gamma}(a, b)$





$$E(\mu|a, b) = \frac{a}{b} = \frac{70}{1} = 70$$

$$Var(\mu|a, b) = \frac{a}{b^2} = \frac{70}{1^2} = 70$$

Quantiles:

```
qgamma(0.025, 700, 10)
```

```
## [1] 64.90978
```

```
qgamma(0.975, 700, 10)
```

```
## [1] 75.27963
```

## Figuring out the Prior:

“My sandwich shops sells about  $50 \pm 2$  sandwiches daily”.

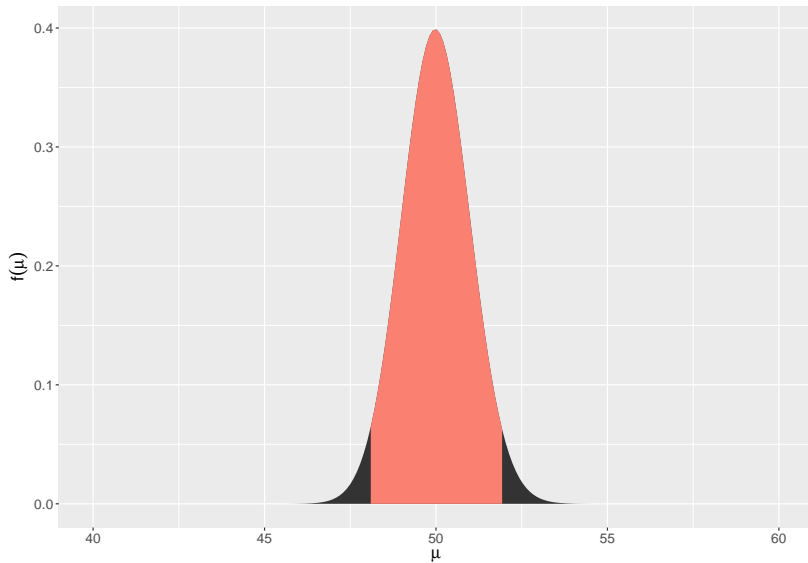
The width of the interval which supports 95% of the probability mass is  $\approx 4sd$ .

Find  $a$  and  $b$  such that

$$\frac{a}{b} = 50 \quad \text{and} \quad \frac{a}{b^2} = 1.$$

I.e.,  $a = 2500$  and  $b = 50$ .

## Plotting the Prior (Just to Check).





## Back to sandwiches: Posterior distribution

$$\mu|x, a, b \sim \text{Gamma}(a + \sum_i x_i, b + n),$$

i.e.

$$\mu|x, a, b \sim \text{Gamma}(1020, 15).$$

## Summarizing the posterior distribution:

**posterior mean and variance:**

$$E(\mu|y, a, b) = \frac{a}{b} = \frac{1020}{15} = 68.$$

$$\text{Var}(\mu|y, a, b) = \frac{a}{b^2} = \frac{1020}{15^2} \approx 4.53 \approx 2.13^2.$$

And **posterior 95% credible interval**: which supports 95% of the posterior density i.e. contains the true value of the parameter w.p. 0.95. (contrast with the classical definition).

```
qgamma(c(0.025, .975), 1020, 15);
```

```
## [1] 63.89042 72.23586
```

## Writing the report:

**Based on our analysis, the posterior expected mean daily demand is 68 sandwiches (95% CI : 63.8, 72.2).**

*Note, that if you were doing it the classical way, employing the Central Limit Theorem and the assumption of Poisson distribution ( $E(x) = \text{Var}(x) = \mu$ ), you would obtain the mean demand estimate of  $\hat{\mu} = \bar{x} = 64$  (s.e. =  $\sqrt{\bar{x}/n} \approx 3.6$ ), and the 95% confidence interval of  $64 \pm 1.96 * 3.6$ , i.e., (57.0, 71.00).*

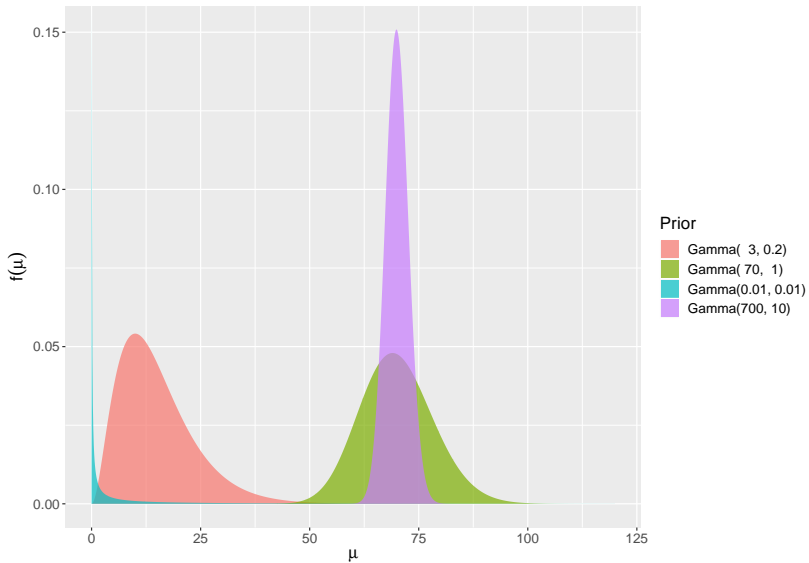
## Summary (so far):

- ▶ The object of Bayesian inference is the posterior distribution of the parameter of interest.
- ▶ The posterior distribution is obtained by combining the data-generating mechanism (likelihood) with the prior information.
- ▶ The posterior distribution can be updated continuously, one observation or one batch of observations at a time.
- ▶ The posterior distribution can be summarised via posterior mean, variance and the Bayesian 95% Credible Interval.
- ▶ The final result will depend on the prior to some extent.

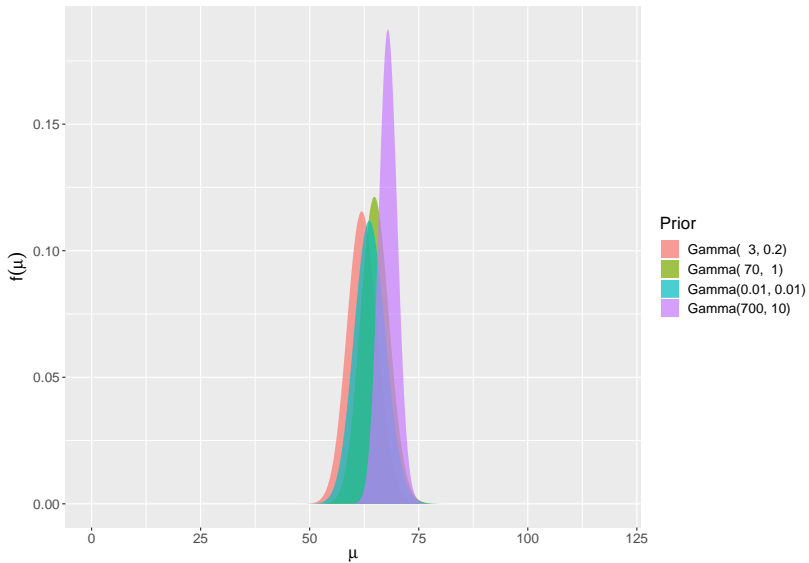
Is Bayesian statistics subjective ???

... what about classical statistics?

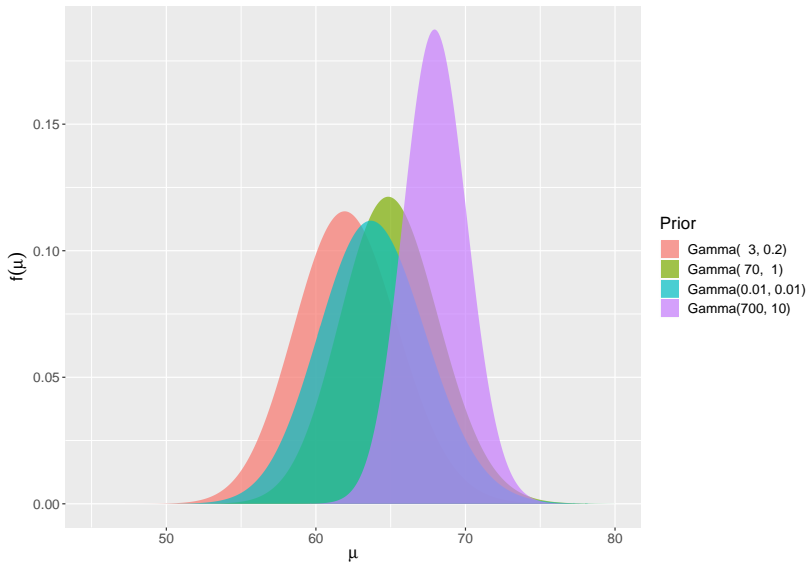
## Different priors ...



## ... Different posteriors



## ... Different posteriors -2





## ... Different posteriors - 3

Prior	a	b	P. Mean	P. Var	2.5%	97.5%
Gamma( 3, 0.2)	3	0.02	64.34	12.8173	57.52	71.55
Gamma( 70, 1)	70	1.00	65.00	10.8333	58.71	71.61
Gamma(0.01, 0.01)	0.01	0.01	63.87	12.7494	57.07	71.06
Gamma(700, 10)	700	10.00	68.00	4.5333	63.89	72.24

## Another look at the posterior distribution:

$$\mu|\mathbf{x}, a, b \sim \text{Gamma}(a + \sum_i x_i, b + n).$$

So, the posterior mean is

$$E(\mu|\mathbf{x}, a, b) = \frac{a + \sum_i x_i}{b + n} = \frac{a + n\bar{x}}{b + n}.$$

As more and more data come in, and the sample size becomes larger and larger,

$$\lim_{n \rightarrow \infty} E(\mu|\mathbf{x}, a, b) = \lim_{n \rightarrow \infty} \frac{a + n\bar{x}}{b + n} = \bar{x}.$$

So...

- ▶ there is a bit of subjectivity in any data analysis
- ▶ the choice of priors matters
- ▶ but it matters less when there is a lot of data.

## A restrictive prior:

Recall, Bayes' Formula:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} \propto f(x|\theta)f(\theta).$$

and consider the following prior distribution function:

$$f(\theta) = \begin{cases} 1/(b-a) & \text{for } \theta \in (a, b) \\ 0 & \text{otherwise,} \end{cases}$$

for some  $a < b$ .

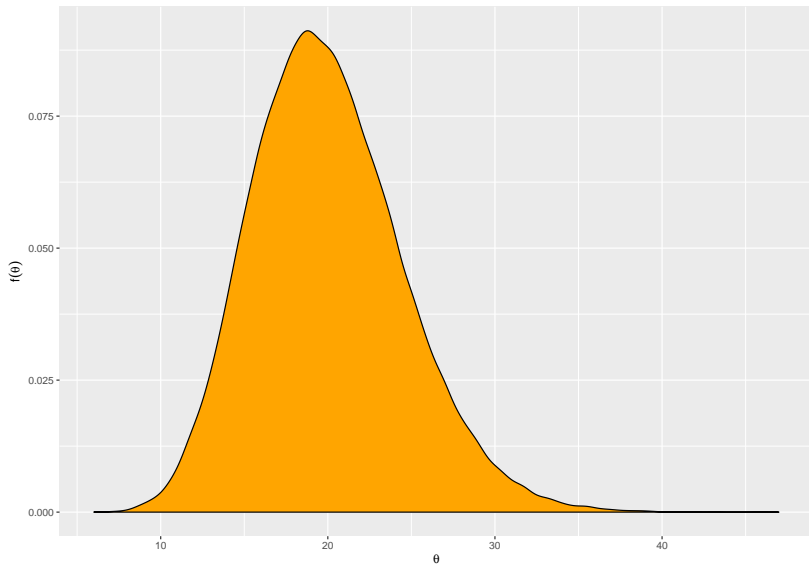
A limited posterior:

$$f(\theta|x) \propto \begin{cases} f(x|\theta) & \text{for } \theta \in (a, b) \\ 0 & \text{otherwise .} \end{cases}$$

## Central Credible Interval vs. Highest Density Region.

- ▶ The central credible interval is delimited by the 2.5%<sup>th</sup> quantile from below and the 97.5%<sup>th</sup> quantile from above.
- ▶ The highest density interval is the smallest interval (or a union of intervals with the shortest summary length) supporting 95% of the probability mass.

Gamma(20,1)

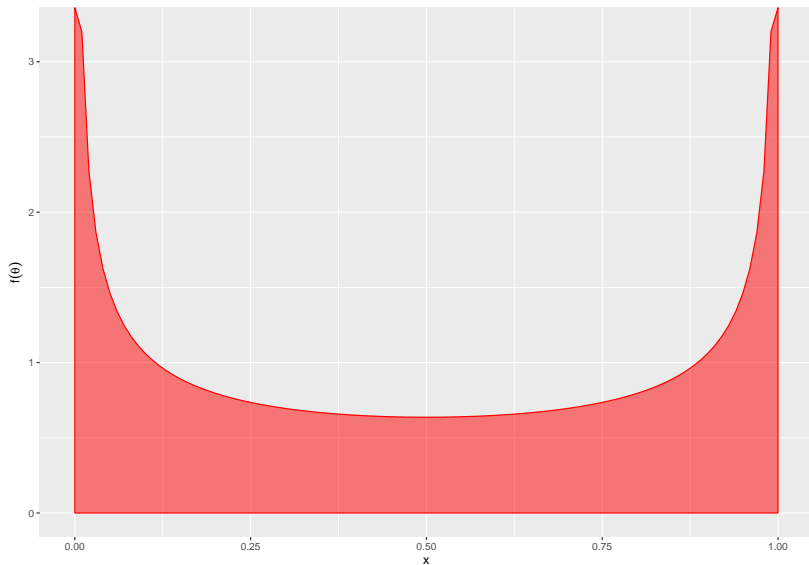


## Gamma(20,1)

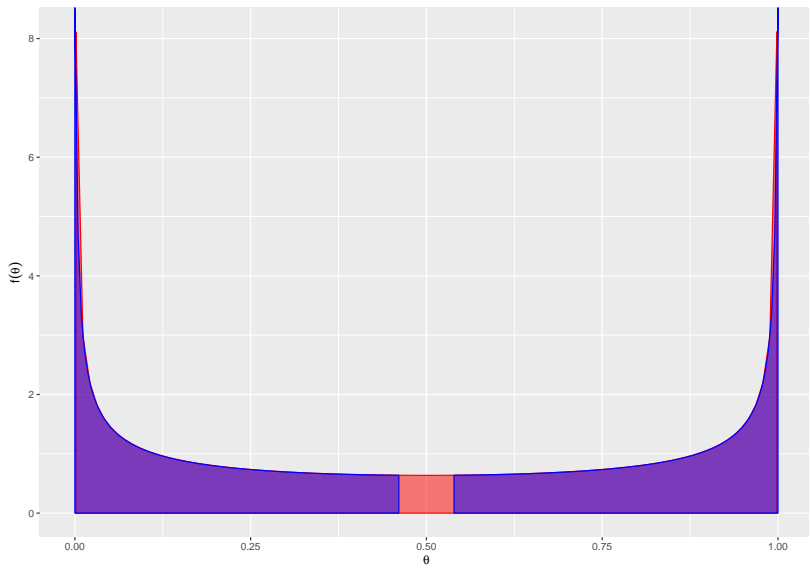
- ▶ 95% (C)CI:  $29.67 - 12.22 = 17.45$ .
- ▶ An interval between the 1.4% quantile and 96.4% quantile:  $28.75 - 11.47 = 17.28$ .
- ▶ The narrowest 95% interval between the 1.63% quantile and 96.63% quantile:  $28.92 - 11.65 = 17.26$ .



Beta(0.5,0.5) - 1



Beta(0.5,0.5) - 2



# Summary so far. . .

- ▶ There are many choices to be made in data analysis. They are subjective to some extent, but it's OK as long as these choices are clearly presented and explained.
- ▶ In general, the more data you have, the less will the choice of priors matter. In many cases, the larger is your sample, the closer will your posterior mean be to the classical maximum likelihood estimator.
- ▶ You can use *non-informative* (also called *vague* or *weak*) prior distribution if you do not have much information about the phenomenon *a priori*.
- ▶ You can use *strong* priors, if you have a good reason for it and explain it well.
- ▶ The *prior* distribution has to be defined *before* you collect the data. Don't use your data to define your prior!
- ▶ Sensitivity analysis can be performed to see how influential the choice of prior is.
- ▶ There are two types of credible intervals: *central* and *highest density*. For unimodal symmetric distributions, they will be one and the same. Central CIs are often easier to compute and they are commonly used by default.

## Posterior predictive distribution

Let  $\tilde{x}$  denote a new, as yet unobserved, datum (demand for sandwiches tomorrow, say). Then, given the data  $\mathbf{x} = \{x_1, \dots, x_n\}$  observed so far,

$$f(\tilde{x}|\mathbf{x}) = \int f(\tilde{x}|\mu)f(\mu|\mathbf{x})d\mu.$$

The distribution of  $\tilde{x}$  given the data  $\mathbf{x}$  is called the *posterior predictive distribution*. You are predicting the future observation after observing some data.

Note that this is different from the posterior distribution of the parameter  $\mu$ .

## Hypothesis Testing:

Is the average demand on Tuesday higher than the average demand on Monday?

## Day-specific posterior distribution:

Table 3: Sandwich Demand

Day	x	post.a	post.b	post.mean	post.lo	post.hi
Monday	50	750	11	68.2	63.4	73.1
Tuesday	65	765	11	69.5	64.7	74.6
Wednesday	72	772	11	70.2	65.3	75.2
Thursday	63	763	11	69.4	64.5	74.4
Friday	70	770	11	70.0	65.1	75.0

## Answering the research questions:

Based on the observations so far, which day of the week has the highest average demand?

# Summary so far

- ▶ It is a common mistake to think about statistics in terms of applying method to the data.
- ▶ Instead, you need to think about answering the actual research question. (And figuring out what data and what methods would get you there.)
- ▶ Posterior distribution concerns the parameter of interest, such as expected Poisson intensity  $\mu|x$ . Posterior predictive distribution concerns the new observations  $\tilde{x}|x$ . You may think of expected demand as a long-term average, and a new observation as demand on any given day. The latter will be much more variable.
- ▶ Simulations and manipulations of samples from posterior distribution are an important part of making Bayesian statistics truly versatile and flexible.
- ▶ It is customary to derive posterior Bayesian probabilities to support various statements. Sometimes, they are called Bayesian  $P$ -values. Keep in mind, however, that the interpretation of Bayesian posterior probabilities is completely different from the interpretation of the frequentist  $p$ -values.
- ▶ In particular, in Bayesian inference, higher posterior probabilities mean that the statement is better supported by the data. Whereas in frequentist inference,  $p < 0.05$  is usually taken as the evidence of strong support for the statement.