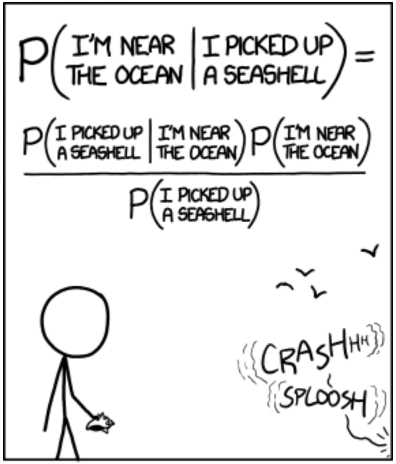
Inferential Statistics



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

comillas.edu

Contents

Recap

Contingency table

ADOLESCENTS' UNDERSTANDING OF SOCIAL CLASS

study examining teens' beliefs about social class

sample: 48 working class and 50 upper middle class 16-year-olds

study design:

- "objective" assignment to social class based on self-reported measures of both parents' occupation and education, and household income
- "subjective" association based on survey questions

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Recap

Marginal Probabilities

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Marginal probabilities are probabilities that can be calculated using the margins of the contingency table.

- What is the probability that a student's objective social class position is upper middle class? → $P(obj\ UMC) = \frac{50}{98} \approx 0.51$

Recap

Joint Probabilities

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Joint probabilities are probabilities that can be calculated using data which are at the intersection of the two events at the contingency table.

- What is the probability that a student's objective position and subjective identity are both upper middle class?→
$P(obj\ UMC \cap sub\ UMC) = \frac{37}{98} \approx 0.38$

Recap
Conditional Probabilities

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Conditional probabilities are joint probabilities where instead of the total contingency only a subgroup of the population is considered given a known condition.

- What is the probability that a student who is objectively in the working class associates with upper middle class?→
  $P(sub\ UMC \mid obj\ WC) = \frac{8}{48} \approx 0.17$

Recap

Bayes Theorem

results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Formally, conditional probabilities are calculated using the Bayes theorem:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- What is the probability that a student who is objectively in the working class associates with upper middle class?→

$$P(sub\ UMC \mid obj\ WC) = \frac{8/98}{48/98} \approx 0.17$$

Recap

Example

Determine if **a person has HIV**. Applicants for the army are screened with an enzyme linked immune absorbent SA, which is commonly referred to as **an ELISA**. **If the sample tested positive** then **one more round of the same ELISA is** performed. **If that test yielded a positive** result, **then a Western Blot** assay, that is more cumbersome to conduct, but has higher accuracy was performed. Only if both of those tests are positive, do the military determine the recruit to have an HIV infection. Based on previous studies:

ELISA:

- Sensitivity (true positive): 93 %
- Specificity (true negative): 99%

Western blot

- Sensitivity: 99.9%
- Specificity: 99.1%

Prevalence of HIV on americans: 1.48/1000

Example: prior and posterior

HIV

$P(HIV) = 0.00148$

- $Elisa +$: $P(+|HIV) = 0.93 \longrightarrow P(+\cap HIV) = 0.0013764$
- $Elisa -$: $P(-|HIV) = 0.07$

$no\ HIV$

$P(no\ HIV) = 0.99852$

- $Elisa +$: $P(+|no\ HIV) = 0.01 \longrightarrow P(+\cap no\ HIV) = 0.0099852$
- $Elisa -$: $P(-|no\ HIV) = 0.99$

$$P(HIV|+) = \frac{P(+\cap HIV)}{P(+)} = \frac{0.0013764}{0.0013764 + 0.0099852} \approx 0.12$$

Example: Bayes updating

HIV → $P(HIV) = 0.12$

- $Elisa\ +$ → $P(+|HIV) = 0.93 \longrightarrow P(+\cap HIV) = 0.1116$
- $Elisa\ -$ → $P(-|HIV) = 0.07$

$no\ HIV$ → $P(no\ HIV) = 0.88$

- $Elisa\ +$ → $P(+|no\ HIV) = 0.01 \longrightarrow P(+\cap no\ HIV) = 0.0088$
- $Elisa\ -$ → $P(-|no\ HIV) = 0.99$

$$P(HIV|+) = \frac{P(+\cap HIV)}{P(+)} = \frac{0.1116}{0.1116 + 0.0088} \approx 0.93$$

Recap

Example: Bayes updating

HIV

$P(HIV) = 0.93$

$WB +$ $P(+|HIV) = 0.9999 \longrightarrow P(+\cap HIV) = 0.929907$

$WB -$ $P(-|HIV) = 0.0001$

$no\ HIV$

$P(no\ HIV) = 0.07$

$WB +$ $P(+|no\ HIV) = 0.009 \longrightarrow P(+\cap no\ HIV) = 0.00063$

$WB -$ $P(-|no\ HIV) = 0.991$

$$P(HIV|+) = \frac{P(+\cap HIV)}{P(+)} = \frac{0.929907}{0.929907 + 0.00063} \approx 0.9994$$

Example

- **Research question**: Is RU-486 an effective “morning after” contraceptive?
- **Participants**: 40 women who came to a health clinic asking for emergency contraception
- **Design**: Random assignment to RU-486 or standard therapy (20 in each group).
- **Data**:
    - 4 out of 20 in RU-486 (treatment group) became pregnant.
    - 16 out of 20 in standard therapy (control group) became pregnant.
- **Question**: How strongly do these data indicate that the treatment is more effective than the control?

For now, let’s focus on binomial random variables (1=success, 0=fail). To calculate the probability of getting k successes in n samples with a probability p of success, use the binomial distribution:

$$P(X = k|\, n, p) = \binom{n}{p} \cdot p^k \cdot (1-p)^{n-k}$$

Inference for a proportion

Frequentist approach

Let's simplify to inference for a single proportion (instead of difference of two proportions) assuming that there is the same probability for a woman to come from the treatment or the control group.

p: probability that a given pregnancy comes from the treatment group

- **Hypothesis**:

$H_0: p = 0.5$ Pregnancy is equally likely from treatment or control group

$H_A: p < 0.5$ Pregnancy is less likely to come from the treatment group.

- **Data**:

$$\hat{p} = \frac{4}{20} = 0.2; SE = \sqrt{\frac{p \cdot (1-p)}{n}} = \sqrt{\frac{0.5 \cdot 0.5}{20}} \approx 0.112$$

- **Z-score**:

$$Z = \frac{0.2 \; - 0.5}{SE} = \frac{-0.3}{0.112} = -2.683$$

- **P-value:**

$$p - value: P(Z < -2.683) \approx 0.0036$$

Inference for a proportion

Bayesian approach

1. Start by setting the **hypothesis** or the **models**.

   These models shall represent the **valid output space**. Although continuous, for simplicity, let's state that p can only take values multiples of 10%, i.e.:

   $$p \in \{10\%, 20\%, 30\%, \dots, 90\%\}$$

   Note that here we are taken **9 models** instead of 1 (frequentist approach).

2. Specify the **prior probabilities**.

   At this point, incorporate information learned from all relevant research up to the current point in time, but not incorporate information from the current experiment. As an example:

Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior	0.06	0.06	0.06	0.06	0.52	0.06	0.06	0.06	0.06	1

Bayesian approach

3. Calculate $P(data|model)$ for each model considered, i.e., **likelihood** of data to happen given each model.

$$P(data|model) = P(k = 4 \mid n = 20, p)$$

In this case, we can use a binomial distribution to calculate this probability.

Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior	0.06	0.06	0.06	0.06	0.52	0.06	0.06	0.06	0.06	1
Likeli.	0.0898	0.2182	0.1304	0.035	0.0046	0.0003	0	0	0	

4. Use Bayes' rule to calculate the posterior probability.

$$P(model|data) = \frac{P(model \cap data)}{P(data)} = \frac{P(data|model) \times P(model)}{P(data)}$$

$$P(model_i|data) = \frac{Prior_i \times likelihood_i}{\sum_{j=1}^{k} Prior_j \times likelihood_j}$$

Inference for a proportion

Bayesian approach

4. Use Bayes' rule to calculate the posterior probability.

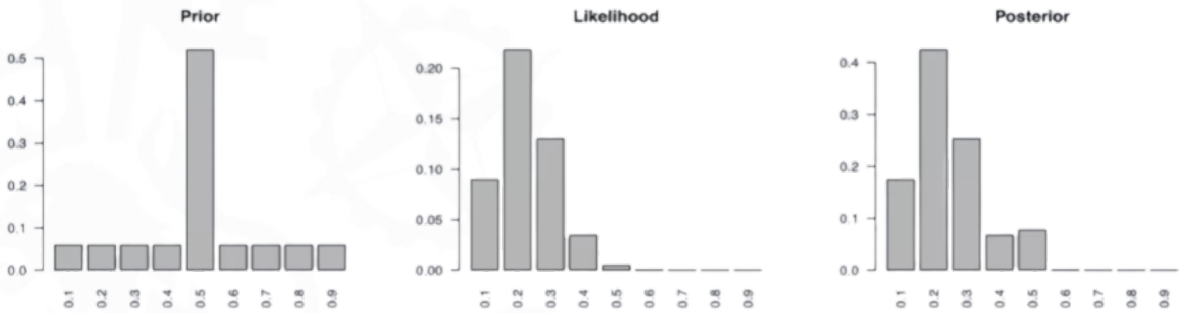
$$P(model_i|data) = \frac{Prior_i \times likelihood_i}{\sum_{j=1}^{k} Prior_j \times likelihood_j}$$

Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior	0.06	0.06	0.06	0.06	0.52	0.06	0.06	0.06	0.06	1
Likeli.	0.0898	0.2182	0.1304	0.035	0.0046	0.0003	0	0	0	
Prior x Likeli.	0.0054	0.0131	0.0078	0.0021	0.0024	0	0	0	0	0.0308
Posterior	0.1748	0.4248	0.2539	0.0681	0.0780	0.0005	0	0	0	1

Inference for a proportion
Bayesian approach

5. Use Bayes' rule to calculate the posterior probability.

$$P(model_i|data) = \frac{Prior_i \times likelihood_i}{\sum_{j=1}^{k} Prior_j \times likelihood_j}$$



Bayesian approach

5. Calculate the probability of $p < 0.5$ using the posterior probabilities.

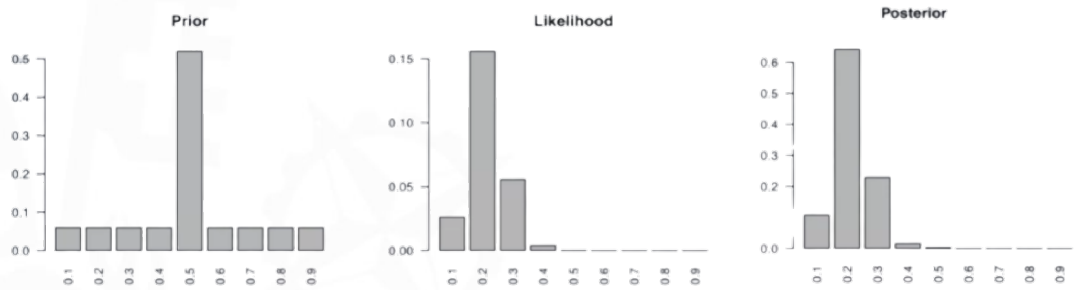
Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Posterior	0.1748	0.4248	0.2539	0.0681	0.0780	0.0005	0	0	0	1

$$P(p < 0.5) = P(p = 0.1) + \cdots + P(p = 0.4) = 0.9216$$

Effect of data size

Imagine we have 40 women with 8 pregnancies in the treatment group:

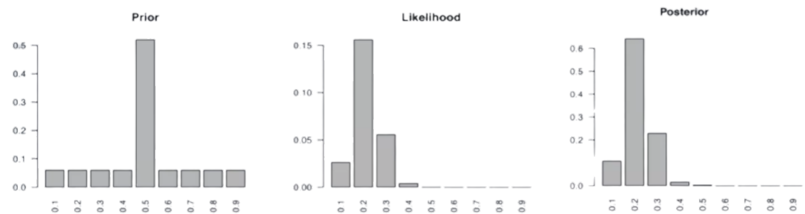
$$k = 8; n = 40$$



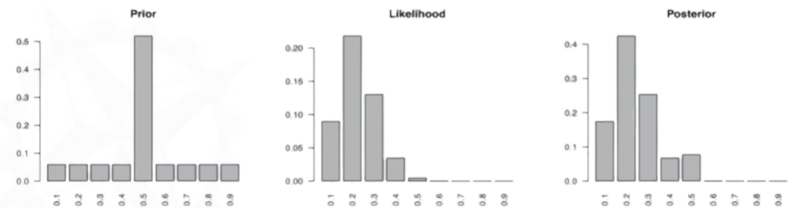
Model corresponding to the sample proportion now have a higher probability

Effect of data size

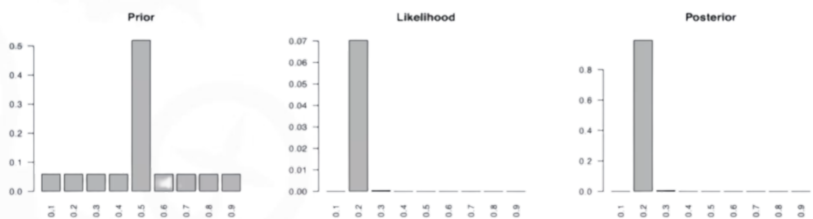
$k = 4; n = 20$



$k = 8; n = 40$



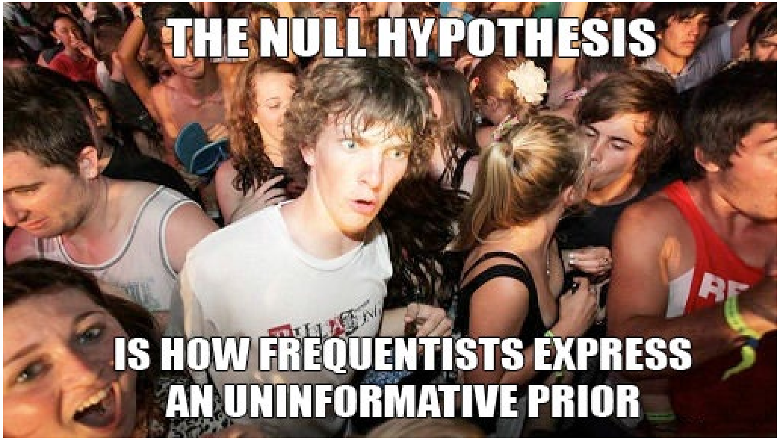
$k = 40; n = 200$



Inference for a proportion
Effect of data size

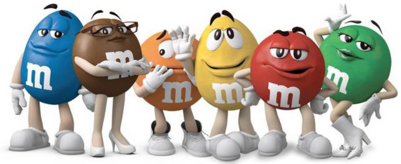
As sample size increases, likelihood dominates prior. This makes sense as it provides more evidence of what is actually happening vs what we believe it happens.

- Even if we have a bad prior, it can be compensated with a large sample size.
- There is one exception, having a model with prior probability equal to 0.



Example

There is a population of M&M's, and in this population the percentage of yellow M&M's is either 10% or 20%. You've been hired as a statistical consultant to decide whether the true percentage of yellow M&M's is 10% or 20%.



Payoffs/Losses:

	True Proportion	
Decision	10%	20%
10%	Bonus	Lose job
20%	Lose job	Bonus

Frequentist vs Bayesian
Example

Data:

- You can buy a random sample from the population
- Each sample cost \$200, and you must buy in \$1000 increments (5 M&Ms at a time)
- You have \$4000 to spend.

For now, we only take five samples and it results in one yellow M&M

Frequentist vs Bayesian

Example

Frequentist approach:

- **Hypothesis**:

$H_0: p = 0.1$

$H_A: p > 0.1$

- **Data**: k=1, n=5
- **P-value:**

$$p - value: P(k \geq 1|n = 5, p = 0.1) = 1 - P(k = 0|n = 5, p = 0.1) \approx 0.41$$

Fail to reject H0, we choose 10% as true proportion with the data we have.

Example

Bayesian approach:

- **Hypothesis (models)**:

$H_1 : p = 0.1$

$H_2 : p = 0.2$

- **Prior probabilities:**

$P(H_1) = 0.5$

$P(H_2) = 0.5$

- **Data**: k=1, n=5

- **Likelihood:** (binomial)

$P(k = 1|H_1) \approx 0.33$

$P(k = 1|H_2) \approx 0.41$

- **Posterior:**

$$P(H_1|k = 1) = \frac{P(H_1) \times P(k = 1|H_1)}{P(k = 1)} = \frac{0.5 \times 0.33}{0.5 \times 0.33 + 0.5 \times 0.41} \approx 0.45$$

$$P(H_2|k = 1) = 1 - 0.45 = 0.55$$

As P(H2) > P(H1) given the observed data, we took 20% as true proportion.

Frequentist vs Bayesian
Example

Increasing the samples, supposing one yellow at a time:

	Frequentist	**Bayesian**	
Obs. Data	P(k or more | 10% yellow)	P(10% yellow | n, k)	P(20% yellow | n, k)
n=5, k=1	0.41	0.45	0.55
n=10, k=2	0.26	0.39	0.61
n=15, k=3	0.18	0.34	0.66
n=20, k=4	0.13	0.29	0.71

Never rejects H0
Highly sensitive to the null hypothesis

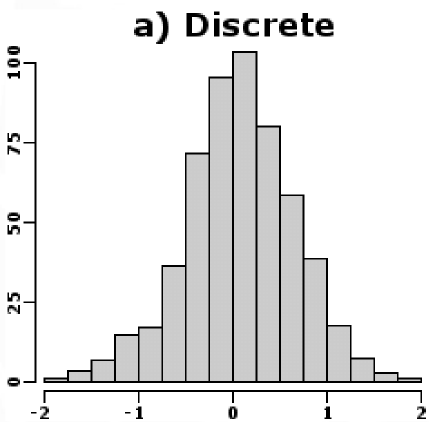
Always accept H2
Information about H1 is also given

Conjugate families

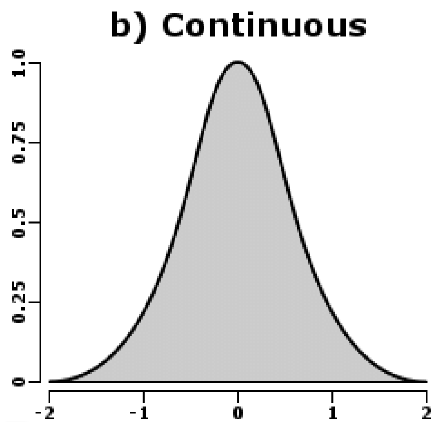
Random variables

Previous examples deal with models belonging to a discrete space, but sometimes we must deal with models belonging to a continuous space:

- Discrete random variables can only take values at separated points.
- Continuous random variables can take any value within an interval.



Probability mass function (pmf)



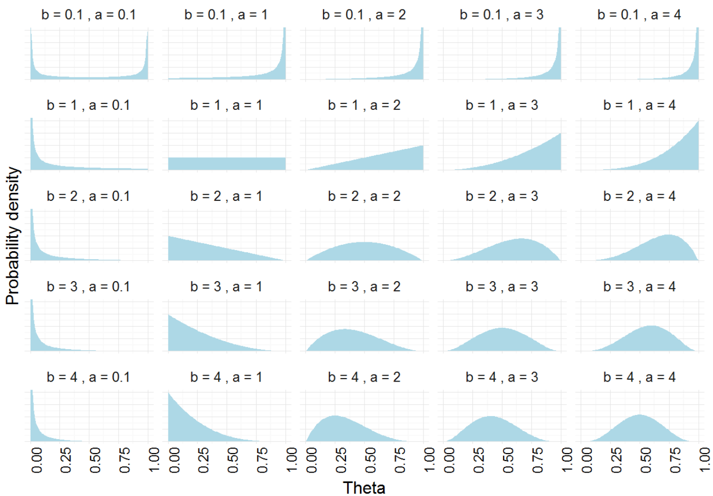
Probability density function (pdf)

Conjugate families

Elicitation

A key process in Bayesian inference is elicitation, i.e., converting previous expert knowledge to form a prior probability distribution of the considered models.

An important pdf that can model several types of prior beliefs for a binary event (including, binomial, normal and uniform) is the beta distribution:



Conjugate families

Beta distribution

Beta family allows to use a continuous pdf to model a discrete pmf.

- It is a continuous probability distribution defined on the interval [0,1].
- Governed by two parameters: **alpha** $\boldsymbol{\alpha}$ and **beta** $\boldsymbol{\beta}$.
- Represents our prior belief about the probability of success for a binary event.
- Its probability density function is given by:

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \qquad \Gamma(n) = (n-1)!$$

Conjugation with beta-binomial family

When using a Beta prior and a Binomial likelihood, the posterior distribution is also a Beta distribution.

The parameters of the posterior Beta distribution are updated as follows:

$$\begin{aligned}\alpha_{posterior} &= \alpha_{prior} + k \\ \beta_{posterior} &= \beta_{prior} + (n-k)\end{aligned}$$

Advantages:

- Analytical tractability: Simplifies Bayesian analysis by providing closed-form solutions for the posterior distribution.
- Easy interpretation: The updated parameters of the posterior distribution directly incorporate the observed data.

In Bayesian inference, **conjugacy** occurs when the **posterior distribution is in the same family as the prior belief but with new parameter values**.

Inference on a Binomial distribution

Research study concerned RU-486 took 800 women, each of whom had intercourse no more than 72 hours before reporting to a family planning clinic to seek contraception. Half of these women were randomly assigned to the standard contraceptive, a large dose of estrogen and progesterone. Other half of the women were assigned RU-486.

Among the RU-486 group, there were no pregnancies. Among those receiving the standard therapy, four became pregnant.

Statistically, these data can be modelled as coming from the binomial distribution. Objective is to determine if RU-486 is more effective than standard procedure.

Conjugate families

Inference on a Binomial distribution

- Prior distribution, as we had no knowledge, can be modelled using the uniform distribution or beta(1, 1).
- Using conjugacy, posterior distribution can be modelled with beta(1+0, 1+4)

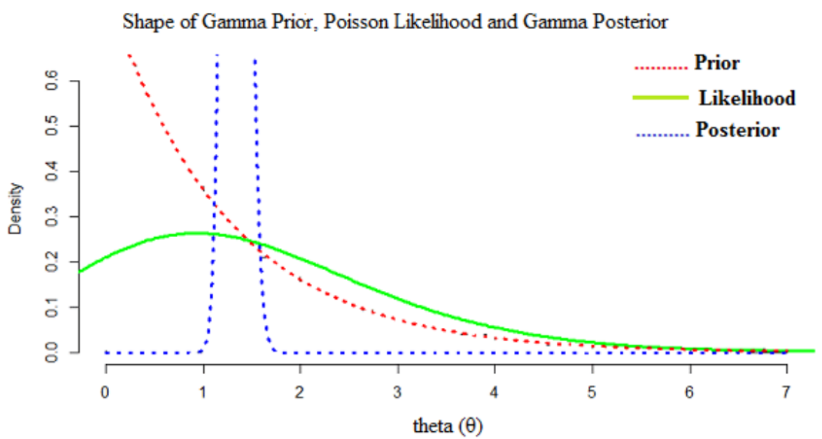
As we know the posterior distribution, statistics of the distribution of p can be obtained easily known alpha and beta:

Mean	$\alpha/(\alpha+\beta)$
Median	$\approx (\alpha - 1/3)/(\alpha+\beta-1/3)$, for $\alpha, \beta > 1$
Mode	$(\alpha - 1)/(\alpha+\beta-2)$, for $\alpha, \beta > 1$
Variance	$\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$
Skewness	$2(\beta-\alpha)\sqrt{\alpha+\beta+1}/\left[(\alpha+\beta+2)\sqrt{\alpha\beta}\right]$
Kurtosis	$6[(\beta-\alpha)^2(\alpha+\beta+1)-\alpha\beta(\alpha+\beta+2)]\ /$ $[\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)]$

The Gamma-Poisson Family

- The Gamma-Poisson conjugate family is a combination of prior and likelihood distributions that simplifies Bayesian inference for count data.

- It involves using a Gamma prior distribution and a Poisson likelihood function, resulting in a Gamma posterior distribution.

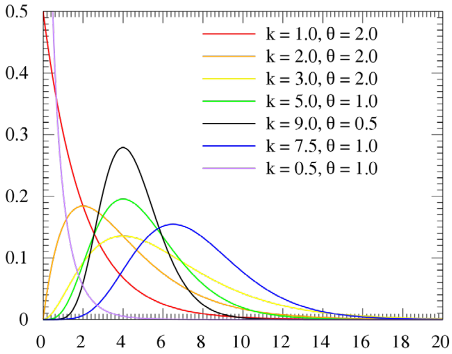


Conjugate families
The Gamma-Poisson Family

Gamma distribution (prior)

- It describes non-negative continuous random variables. Its pdf is defined on the interval (0, ∞)
- It is governed by two parameters: k and θ.
- Represents our prior belief about the rate (λ) of a Poisson-distributed event.
- PDF: $f(\lambda) = \frac{\lambda^{k-1} \cdot e^{\frac{-\lambda}{\theta}}}{\theta^k \cdot \Gamma(k)}$ where $\Gamma(k)$ is the Gamma function

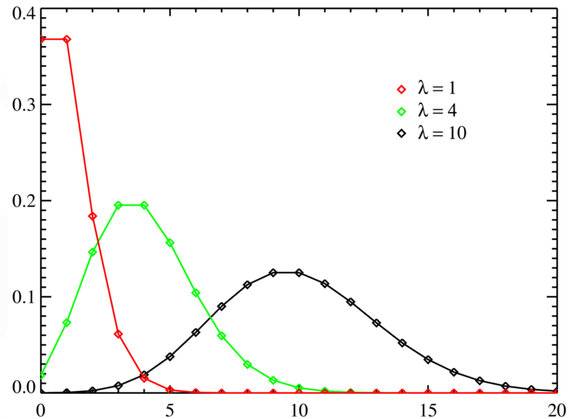
If a random variable is being modelled by a gamma function:
Mean: $\bar{x} = \lambda = k \cdot \theta$
Std. Deviation: $\sigma = uncertainty(\lambda) = \theta \cdot \sqrt{k}$



The Gamma-Poisson Family

Poisson distribution (Likelihood)

- It models non-negative discrete numbers, usually the probability of an event happening a defined number of times in a time range.
- It is governed by a single parameter: rate (λ)
- It is suitable for modelling events occurring independently and a constant average rate.
- PMF: $P(X = k) = \frac{\lambda^k . e^{(-\lambda)}}{k!}$



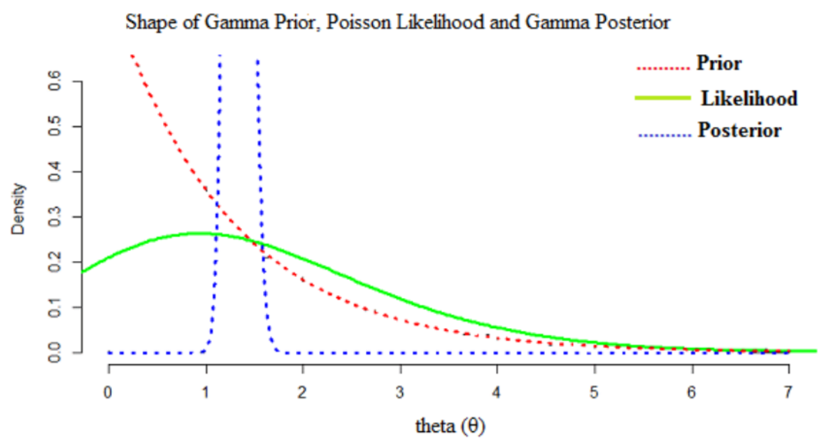
The Gamma-Poisson Family

Conjugation

- When using a Gamma prior and a Poisson likelihood, the posterior distribution is also a Gamma distribution.
- The parameters of the posterior Gamma distribution are updated as follows:

$$k_{posterior} = k_{prior} + \sum x_i$$

$$\theta_{posterior} = \frac{\theta_{prior}}{n_{x_i} \cdot \theta_{prior} + 1}$$



The Gamma-Poisson Family

Example

You are a general, and you believe your soldiers casualties happens following a Poisson distribution with a rate of 0.75 soldiers per year per soldier unit. The uncertainty about this rate is around 1.
You had data of 15 soldier units for the last 20 years, and you observe that 200 soldiers died during those 20 years. For this data, using the gamma-poisson distribution, we want to get a better estimation of the rate of soldiers casualty per year per unit.

$$\lambda_{prior} = 0.75 = k \cdot \theta$$
$$\sigma = 1 = \theta\sqrt{k}$$

The Gamma-Poisson Family

Example

You are a general, and you believe your soldiers casualties happens following a Poisson distribution with a rate of 0.75 soldiers year per soldier per unit. The uncertainty about this rate is around 1.
You had data of 15 soldier units for the last 20 years, and you observe that 200 soldiers died during those 20 years. For this data, using the gamma-poisson distribution, we want to get a better estimation of the rate of soldiers casualty per year per unit.

$\lambda_{prior} = 0.75 = k \cdot \theta$

$\sigma = 1 = \theta\sqrt{k}$

$k = \frac{9}{16}$

$\theta = \frac{4}{3}$

The Gamma-Poisson Family

Example

You are a general, and you believe your soldiers casualties happens following a Poisson distribution with a rate of 0.75 soldiers year per soldier per unit. The uncertainty about this rate is around 1.

You had data of 15 soldier units for the last 20 years, and you observe that 200 soldiers died during those 20 years. For this data, using the gamma-poisson distribution, we want to get a better estimation of the rate of soldiers casualty per year per unit.

$$\lambda_{prior} = 0.75 = k \cdot \theta$$

$$\sigma = 1 = \theta\sqrt{k}$$

$$k = \frac{9}{16}$$

$$\theta = \frac{4}{3}$$

Data:

$$n = 20 \cdot 15 = 300$$

$$\sum x_i = 200$$

The Gamma-Poisson Family

Example

You are a general, and you believe your soldiers casualties happens following a Poisson distribution with a rate of 0.75 soldiers year per soldier per unit. The uncertainty about this rate is around 1.
You had data of 15 soldier units for the last 20 years, and you observe that 200 soldiers died during those 20 years. For this data, using the gamma-poisson distribution, we want to get a better estimation of the rate of soldiers casualty per year per unit.

$$\lambda_{prior} = 0.75 = k \cdot \theta$$
$$\sigma = 1 = \theta\sqrt{k}$$
$$k = \frac{9}{16}$$
$$\theta = \frac{4}{3}$$

Data:

$$n = 20 \cdot 15 = 300$$
$$\sum x_i = 200$$

Update:

$$k_{posterior} = \frac{9}{16} + 200 = 200.5625$$
$$\theta_{posterior} = \frac{4/3}{300 \cdot 4/3 + 1} = 0.0033$$

The Gamma-Poisson Family

Example

You are a general, and you believe your soldiers casualties happens following a Poisson distribution with a rate of 0.75 soldiers year per soldier per unit. The uncertainty about this rate is around 1.
You had data of 15 soldier units for the last 20 years, and you observe that 200 soldiers died during those 20 years. For this data, using the gamma-poisson distribution, we want to get a better estimation of the rate of soldiers casualty per year per unit.

$$\lambda_{prior} = 0.75 = k \cdot \theta$$

$$\sigma = 1 = \theta\sqrt{k}$$

$$k = \frac{9}{16}$$

$$\theta = \frac{4}{3}$$

Data:

$$n = 20 \cdot 15 = 300$$

$$\sum x_i = 200$$

Update:

$$k_{posterior} = \frac{9}{16} + 200 = 200.5625$$

$$\theta_{posterior} = \frac{4/3}{300 \cdot 4/3 + 1} = 0.0033$$

Final

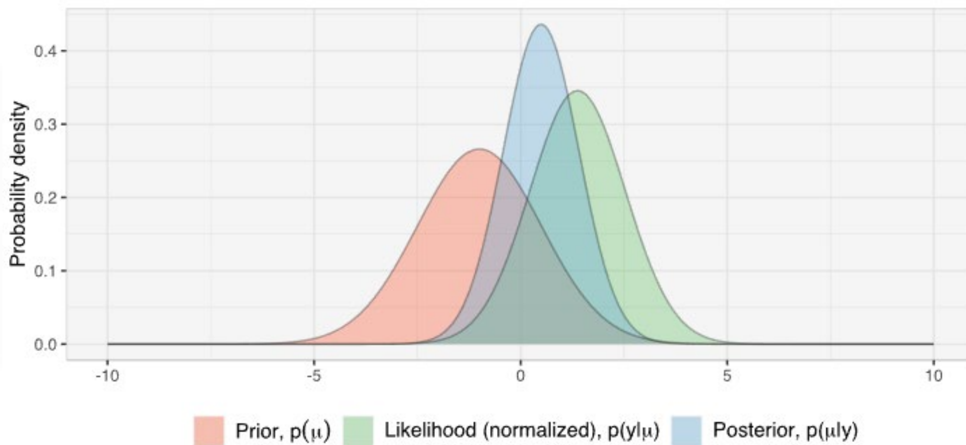
$$\lambda_{posterior} = 200.5625 \cdot 0.0033 = 0.67$$

$$Uncertainty = 0.0033 \cdot \sqrt{200.5625} = 0.0047$$

Conjugate families

The Normal-Normal Family

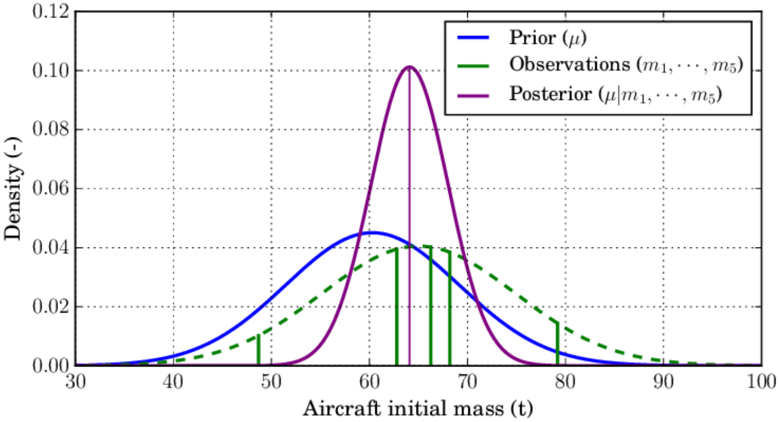
- The Normal-Normal conjugate family is a combination of prior and likelihood distributions that simplifies Bayesian inference for continuous data with normally (with known variance) distributed errors.
- It involves using a Normal prior distribution for the mean and a Normal likelihood function, resulting in a Normal posterior distribution.
- Data samples must be independent and come from a normal distribution with known standard deviation but unknown mean (which we want to infer).



The Normal-Normal Family

Normal distribution (prior)

- A continuous probability distribution defined on the interval (-∞, ∞).
- Governed by two parameters: mean (ν) and variance (τ^2).
- Represents our prior belief about the mean (ν) of a normally distributed random variable.
- PDF: $f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \tau^2}} e^{-\frac{(x-\nu)^2}{2 \cdot \tau^2}}$



Conjugate families

The Normal-Normal Family

Conjugation

- When using a Normal prior and a Normal likelihood, the posterior distribution is also a Normal distribution.
- The parameters of the posterior Normal distribution are updated as follows:

$$\tau^2_{posterior} = \sqrt{\frac{\sigma^2 \cdot \tau^2_{prior}}{\sigma^2 + n \cdot \tau^2_{prior}}} \nu_{posterior} = \frac{\sigma^2 \cdot \nu_{prior} + n \cdot \tau^2_{prior} \cdot \bar{x}}{\sigma^2 + n \cdot \tau^2_{prior}}$$

Conjugate families

The Normal-Normal Family

Example

Suppose an analytical chemist wants to measure the mass of a sample of ammonium nitrate. Her balance has a known standard deviation of 0.2 milligrams. By looking at the sample, she thinks this mass is about 10 milligrams and based on her previous experience in estimating masses her uncertainty in that guess of her prior is 2. So, she decides that her prior for the mass of the sample is a normal distribution with mean, 10 milligrams, and standard deviation, 2 milligrams.

$$x_{prior} \sim N(\nu_{prior} = 10, \tau_{prior} = 2)$$

Now she collects five measurements on the sample and finds that the average of those is 10.5.

$$\bar{x} = 10.5$$

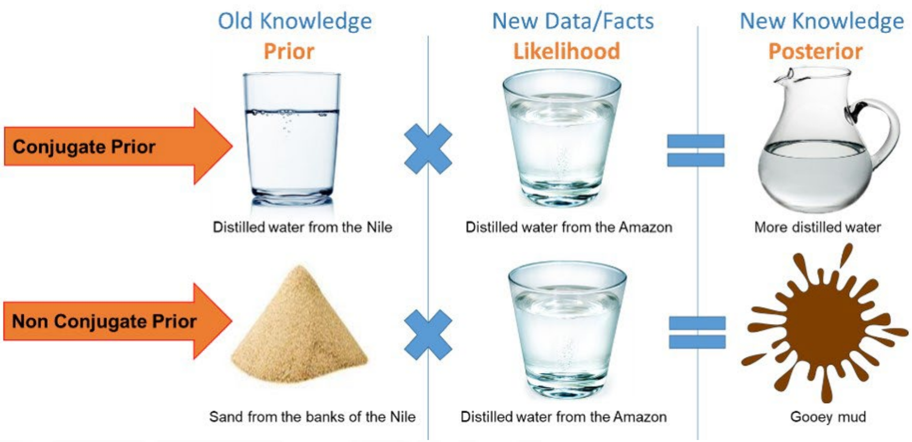
By conjugacy of the normal-normal family, our posterior belief about the mass of the sample has the normal distribution. Using the data:

$$x_{posterior} \sim N(\nu_{posterior} = 10.499, \tau_{posterior} = 0.089)$$

Non-Conjugate prior
The issue

Sometimes, using a conjugate prior is not an option:

- There is a previous reference prior (from a previous study), which injects the minimum amount of personal belief into the analysis
- The personal belief about the problem cannot be expressed in terms of a convenient conjugate prior.



Non-Conjugate prior
Example

For the RU-486 case, suppose previous studies stated that the new anticonceptive in no way can be worse than the standard procedure. The prior of our study would change from something like:

Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior	0.06	0.06	0.06	0.06	0.52	0.06	0.06	0.06	0.06	1

To something like:

Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior	0.12	0.12	0.12	0.12	0.52	0	0	0	0	1

And there is no conjugate prior that represent this data (or the continuous equivalent).

Non-Conjugate prior

Use a cannon to kill the fly

There is no simple way to compute p based on the prior of the previous example. Until recent times, there were no resources to do so.

Now, we have JAGS (Just Another Gibbs Sampler)

Just
Another
Gibbs
Sampler

Non-Conjugate prior
JAGS

JAGS (Just Another Gibbs Sampler) is a powerful software tool for performing Bayesian inference when conjugate priors are not applicable.

- It uses Markov Chain Monte Carlo (MCMC) methods, specifically Gibbs Sampling, to estimate posterior distributions in complex models.
- Handles complex models: JAGS can handle non-conjugate priors and hierarchical models with multiple layers.
- Automatic: JAGS automates the MCMC process, simplifying the implementation of Bayesian inference.
- There are several alternatives to JAGS: OpenBUGS, STAN, …

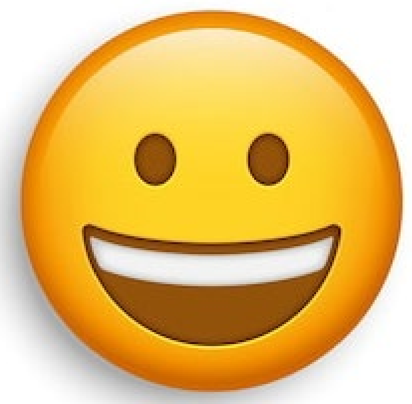
Credible intervals

Interpretation of confidence intervals

There was a lie in this course, the interpretation of confidence intervals.

$$CI: pe \pm cv \cdot se$$

As frequentists, confidence intervals has been described as the *"range from a lower threshold and an upper threshold where the true mean of the population lies with a 95% probability"*.



Interpretation of confidence intervals

There was a lie in this course, the interpretation of confidence intervals.

$$CI: pe \pm cv \cdot se$$

As frequentists, confidence intervals has been described as the "*range from a lower threshold and an upper threshold where the true mean of the population lies with a 95% probability*".

And this is not true, the true mean would be inside that range (probability = 1) or not (probability = 0), and we don't know which is the case.

the interpretation of, for example, a 95% confidence interval on the mean. Is that 95% of similarly constructed confidence intervals will contain the true mean.

Credible intervals

In Bayesian inference, the alternatives to confidence intervals are the **credible intervals** (or **highest density intervals HDI**). These intervals are ranges of values where we think the true mean is contained with a 95% probability.

For a Bayesian, recalling the RU-486 example, a 95% credible Interval is any L and U such that the posterior probability is $L < p < U = 0.95$. As there is an infinity number of intervals, usually the shortest interval is chosen as credible interval.

For the example, posterior distribution was beta(1,5), whose pdf was:

$$f(p) = 5 \cdot (1-p)^4 \; for \; 0 \le p \le 1$$

To obtain the area under the curve, we can use the cumulative distribution function (CDF):

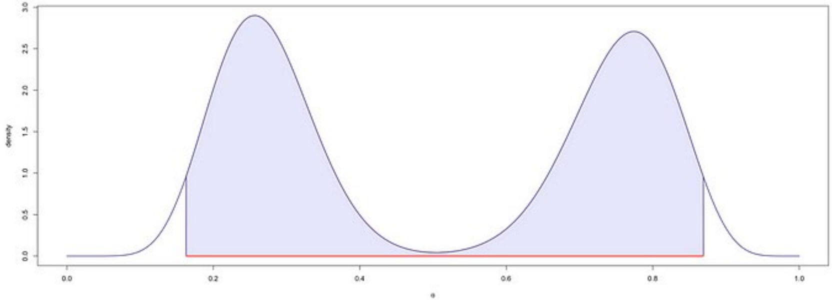
$$F_X(x) = P(X \le x) = 1 - (1-x)^5 \; for \; 0 \le x \le 1$$

And find the shortest interval that meets $F_X(U) - F_X(L) = 0.95$

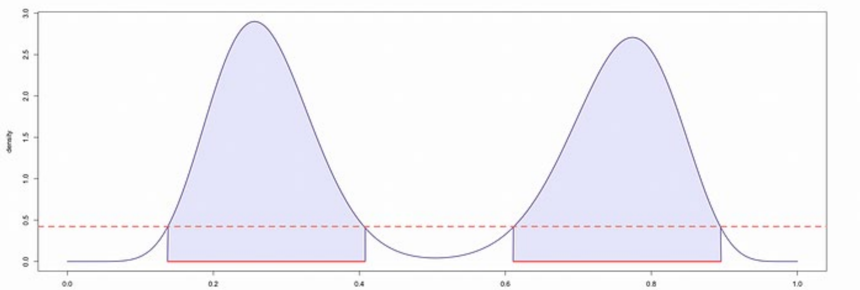
Maths required to calculate this interval requires optimization and finding the inverse of the CDF, so just use a suitable programming package.

Credible intervals

Credible intervals are a much better representation than confidence intervals in multimodal distributions. Using confidence intervals:



However, using credible intervals, we can have **credible sets**:



Predictive inference

Predictive inference arises when the goal is not to find a posterior distribution over some parameter, but rather to find a posterior distribution over some random variable that depends on the parameter.

Specifically, the inference on X is based on the PDF of a random variable X given a parameter θ: $f(x \mid \theta)$, with a prior personal belief on the parameter: $\pi(\theta)$

Usually, this requires integration:

$$CDF: P(X \leq x) = \int_{-\infty}^{x} P(X \leq x \mid \theta)\, \pi(\theta)\, d\theta$$

However, let's check a trivial example.

Example

Suppose you have two coins. One coin has probability 0.7 of coming up heads, and the other has probability 0.4 of coming up heads. You are playing a gambling game with a friend, and you draw one of those two coins at random from a bag.

Before you start the game, your prior belief is that you drew the 0.7 coin is 0.5. That is because both coins were equally likely to be drawn. In this game, you win if the coin comes up heads.

$$p_{prior}^{0.7} = p_{prior}^{0.4} = 0.5$$

Before doing the actual bet, you get a coin from the bag and you throw the coin two times, getting two heads. What is your posterior on the probability that you have drawn the 0.7 coin?

Example

Suppose you have two coins. One coin has probability 0.7 of coming up heads, and the other has probability 0.4 of coming up heads. You are playing a gambling game with a friend, and you draw one of those two coins at random from a bag.

Before you start the game, your prior belief is that you drew the 0.7 coin is 0.5. That is because both coins were equally likely to be drawn. In this game, you win if the coin comes up heads.

$$p_{prior}^{0.7} = p_{prior}^{0.4} = 0.5$$

Before doing the actual bet, you get a coin from the bag and you throw the coin two times, getting two heads. What is your posterior on the probability that you have drawn the 0.7 coin?

$$p_{posterior}^{0.7} = \frac{P(2\ Heads \mid 0.7) \times 0.5}{P(2\ Heads \mid 0.7) \times 0.5 + P(2\ Heads \mid 0.4) \times 0.5} = 0.754$$

Example

But… the bet is not on what coin you draw, but to toss heads when you throw. So, what is the probability of tossing heads in your next throw?

$$P(heads) = P(head|0.7) \times p_{posterior}^{0.7} + P(head|0.4) \times p_{posterior}^{0.4} = 0.7 \times 0.754 + 0.4 \times (1 - 0.754) = 0.626$$

Point estimates

For now, we focus on getting a range of probable values for the variable of interest. But what if we want to predict a unique value?

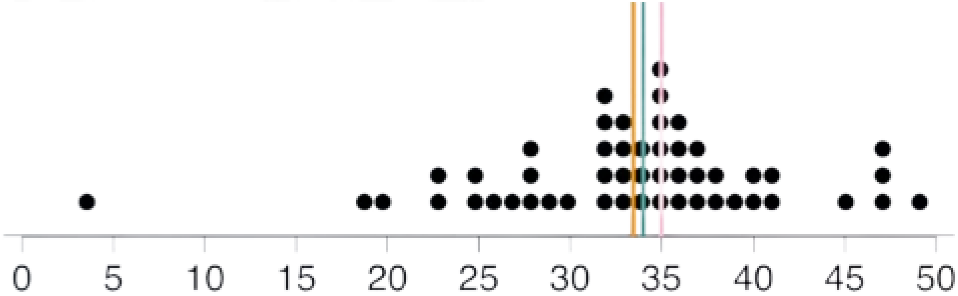
Then, we need decisión theory. In decision theory, one seeks to minimize one's expected loss. Loss is tricky and the best estimate value depends heavily on it:

Loss	**Best estimate**
Linear	median
Squared	mean
0/1	mode

Loss functions

Example

You work at a car dealership. Your boss wants to know how many cars the dealership will sell per month. An analyst who has worked with past data from your company provided you a distribution that shows the probability of number of cars the dealership will sell per month. In Bayesian lingo, this is called the posterior distribution.



Predictive inference

Loss functions

L_0: 0/1 *Loss*

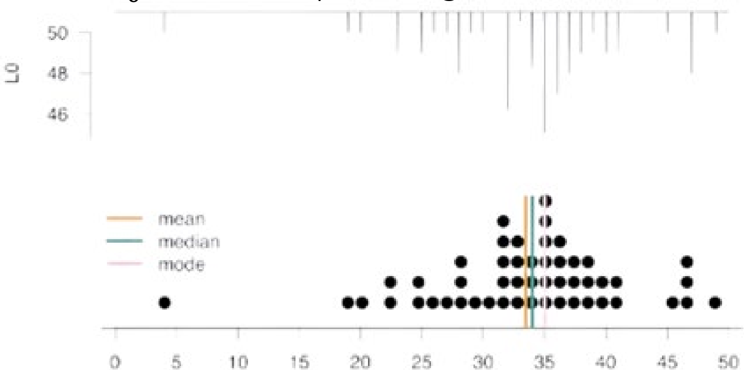
$$L_{0,i}(g) = \begin{cases} 0 & if\ g = x_i \\ 1 & otherwise \end{cases}$$

$$L_0 = \sum L_{0,i}(g)$$

$g = 30$

i	x_i	L_0: 0/1
1	4	1
2	19	1
…	…	…
14	30	0
…	…	…
51	49	1
	Total	50

L_0 when each possible guess is considered



L_0 is minimized if g=mode

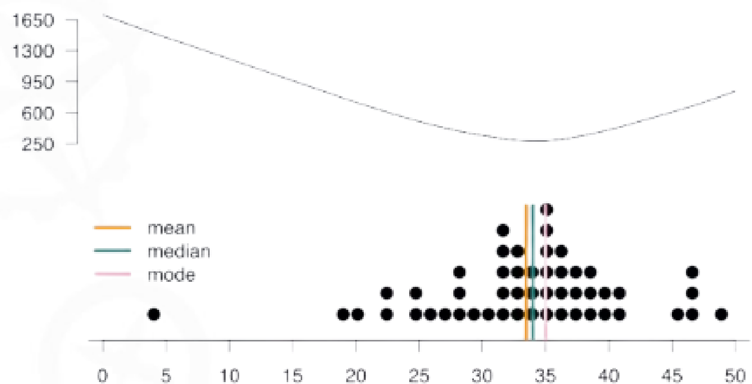
Loss functions

L_1: *linear loss*

$$L_0 = \sum |x_i - g|$$

$g = 30$

i	x_i	$L_1: lin.$
1	4	26
2	19	11
…	…	…
14	30	0
…	…	…
51	49	19
	Total	346



L_1 is minimized if g=median

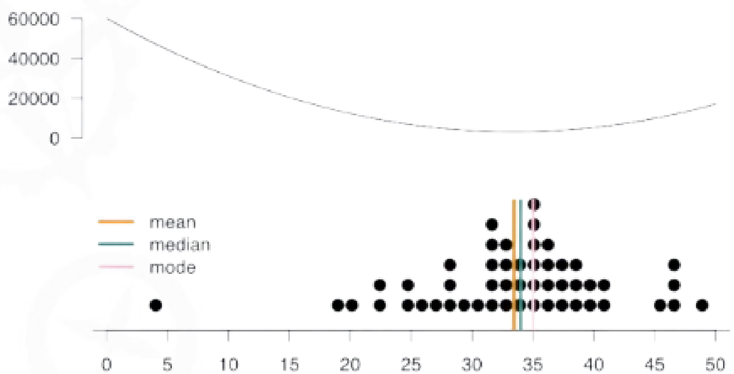
Predictive inference
Loss functions

L_2: square loss

$$L_0 = \sum (x_i - g)^2$$

$g = 30$

i	x_i	$L_1: lin.$
1	4	676
2	19	121
…	…	…
14	30	0
…	…	…
51	49	361
	Total	3732



L_2 is minimized if g=mean

Hypothesis testing

Posterior probabilities & decision making

Suppose you have two competing hypotheses H_1 and H_2.

Then:

- $P(H_1 is\ true\ |\ data)$ = posterior probability of H_1
- $P(H_2\ is\ true\ |\ data)$ = posterior probability of H_2

A potential decision criterion is to choose the hypothesis with the higher posterior probability:

Reject H_1 if $P(H_1 is\ true\ |\ data) < P(H_2 is\ true\ |\ data)$

But we can also choose the hypothesis which minimize a given loss.

Hypothesis testing: example

HIV Testing with ELISA

- H_1 : Patient does not have HIV
- H_2: Patient has HIV

There are two mutually exclusive hypothesis that cover the entire decision space.

Loss function for a given decision can be defined as $L(d)$ where d is the decision we made.

Possible decisions:

- d_1: choose H_1 – decide that patient does not have HIV.
- d_2: choose H_2 – decide that patient has HIV.

Hypothesis testing: example

HIV Testing with ELISA

$d = d_1$

- Right: decide patient does not have HIV and indeed they don't.

$$L(d_1) = 0$$

- Wrong: decide patient does not have HIV, but they do.

$$L(d_1) = w_1$$

$d = d_2$

- Right: decide patient has HIV and indeed they do.

$$L(d_1) = 0$$

- Wrong: decide patient has HIV, but they don't.

$$L(d_2) = w_2$$

Hypothesis testing: example

HIV Testing with ELISA

Hypothesis	H_1: Patient does not have HIV H_2: Patient has HIV	Posterior	$P(H_1 \mid +) \approx 0.88$ $P(H_2 \mid +) \approx 0.12$
Decision	d_1: Decide patient does not have HIV d_2: Decide patient has HIV	Expected losses	
Losses	$L(d_1) = \begin{cases} 0 & if\ d_1\ is\ right \\ w_1 = 1000 & if\ d_1\ is\ wrong \end{cases}$ $L(d_2) = \begin{cases} 0 & if\ d_2\ is\ right \\ w_2 = 10 & if\ d_2\ is\ wrong \end{cases}$		$E[L(d_1)]$ $= 0.88 \cdot 0 + 0.12 \cdot 1000$ $= 120$ $E[L(d_2)]$ $= 0.88 \cdot 10 + 0.12 \cdot 0 = 8.8$

Choose d_2 as it minimizes the expected loss

Predictive inference

Bayes Factors

Prior odds

Ratio of the prior probabilities of hypothesis

$$O[H_1 : H_2] = \frac{P(H_1)}{P(H_2)}$$

Posterior odds

Ratio of the posterior probabilities of hypothesis

$$PO[H_1 : H_2] = \frac{P(H_1 \mid data)}{P(H_2 \mid data)} = \frac{P(data \mid H_1) \cdot P(H_1)/P(data)}{P(data \mid H_1) \cdot P(H_1)/P(data)} =$$

$$= \frac{P(data \mid H_1)}{P(data \mid H_2)} \cdot \frac{P(H_1)}{P(H_2)}$$

The **bayes factor** is defined as the ratio between the posterior and the prior odds.

$$BF[H_1 : H_2] = \frac{P(data \mid H_1)}{P(data \mid H_2)} = \frac{PO[H_1 : H_2]}{O[H_1 : H_2]}$$

Bayes Factors

The **bayes factor** is defined as the ratio between the posterior and the prior odds.

$$BF[H_1 : H_2] = \frac{P(data \mid H_1)}{P(data \mid H_2)}$$

- Quantifies the evidence of data arising from H_1 to H_2.
- Discrete case: ratio of the likelihoods of the observed data under the two hypotheses.
- Continuous case: ratio of the marginal likelihoods

$$BF[H_1 : H_2] = \frac{\int P(data \mid \theta, H_1)\, d\theta}{\int P(data \mid \theta, H_2)\, d\theta}$$

Bayes Factors: example

HIV Testing with ELISA

Hypothesis	H_1: Patient does not have HIV H_2: Patient has HIV	
priors	$P(H_1) = 0.99852$ $P(H_2) = 0.00148$	$O[H_1:H_2] = \frac{P(H_1)}{P(H_2)}$=674.6757
posteriors	$P(H_1 \mid +) = 0.8788551$ $P(H_2 \mid +) = 0.1211449$	$PO[H_1:H_2] = \frac{P(H_1 \mid +)}{P(H_2 \mid +)} = 7.254578$
Bayes factor	$BF[H_1:H_2] = \frac{PO[H_1:H_2]}{O[H_1:H_2]} = \frac{7.254578}{674.6757} \approx 0.0108 \approx \frac{P(+ \mid H_1)}{P(+ \mid H_2)} = \frac{0.01}{0.93}$	

Predictive inference

Bayes Factors

Interpreting the Bayes Factor

Jeffreys (1961)

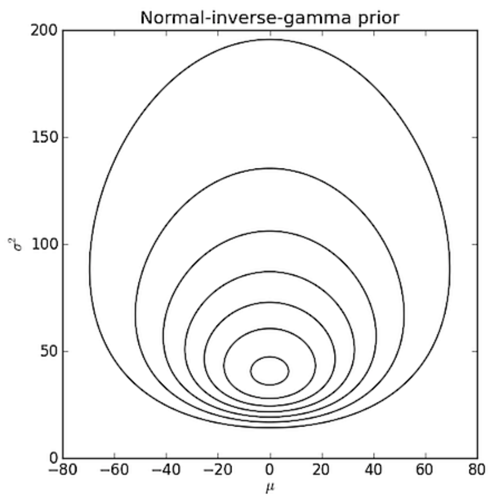
$\mathbf{BF[H_A: H_B]}$	Evidence against $\mathbf{H_B}$
1 to 3	Not worth a bare mention
3 to 20	Positive
20 to 150	Strong
>150	Very strong

$$BF[H_2: H_1] = \frac{1}{BF[H_1: H_2]} = \frac{1}{0.0108} = 92.59259$$

Conjugate families (II)
The Normal-Gamma Family

- The Normal-Gamma conjugate family is a combination of prior and likelihood distributions that simplifies Bayesian inference for continuous data with normal distributed error (with unknown variance).

- The Normal-Gamma distribution is a joint distribution of two random variables: a Gaussian random variable (mean) and a Gamma random variable (precision).



Conjugate families (II)
The Normal-Gamma Family

NormalGamma distribution

- For normal data, the conjugate prior distribution of μ when standard deviation σ is known is a normal distribution:

$$\mu \mid \sigma \sim N(m_0, \sigma^2/n_0)$$

- Since the variance is non-negative, continuous, and with no upper limit, a good candidate for conjugate prior for that parameter could be a Gamma. However, is better for the inverse of the variance, or the **precision**.

$$\phi \sim \mathrm{Gamma}\left(\frac{v_0}{2}, s_0^2 \cdot \frac{v_0}{2}\right); \;\; \phi = 1/\sigma^2$$

- Together:

$$(\mu, \phi) \sim \mathrm{NormalGamma}(m_0, n_0, s_0^2, v_0)$$
$$m_0\text{: prior mean}$$
$$n_0\text{: prior sample size}$$
$$s_0^2\text{: prior variance}$$
$$v_0\text{: prior degrees of variance}$$

The Normal-Gamma Family

Conjugation

- Prior: $(\mu, \phi) \sim \text{NormalGamma}(m_0, n_0, s_0^2, v_0)$
- Posterior: $(\mu, \phi) \mid \text{data} \sim \text{NormalGamma}(m_n, n_n, s_n^2, v_n)$
- The parameters of the posterior NormalGamma distribution are updated as follows:

$$m_n = \frac{n \cdot \bar{Y} + n_0 \cdot m_0}{n + n_0}$$

$$n_n = n_0 + n$$

$$v_n = v_0 + n$$

$$s_n^2 = \frac{1}{v_n} \cdot \left[s_0^2 \cdot v_0 + s^2 \cdot (n-1) + \frac{n_0 \cdot n}{n_n} \cdot (\bar{Y} - m_0)^2\right]$$

The Normal-Gamma Family

Uncertainty of μ

- In this case, the uncertainty around μ is given by a student t distribution:

$$\mu \mid data \sim t(v_n, m_n, s^2/n_n) \leftrightarrow t = \frac{\mu - m_n}{\frac{s_n}{\sqrt{n_n}}} \sim t(v_n, 0, 1)$$

The Normal-Gamma Family

Example

Let's try to infer the total trihalomethane, or TTHM, in tap water from a city in North Carolina. Using prior information about TTHM from the city, as prior we shall use a NormalGamma distribution with a prior mean of 35 parts per billion based on a prior sample size of 25 and the estimate of the variance of 156.25 with degrees of freedom 24. In a new sample of size 28, the mean value of TTHM is 55.5 parts per billion with a sample variance of 540.7

Prior: $\mathrm{NormalGamma}(35, 25, 156.25, 24)$

Data: $\bar{Y} = 55.5, s^2 = 540.7, n = 28$

Update:

$$n_n = 25 + 28 = 53 \qquad m_n = \frac{28 \cdot 55.5 + 25 \cdot 35}{53} = 45.8 \qquad v_n = 24 + 28 = 52$$

$$s_n^2 = \frac{1}{52} \cdot \left[27 \cdot 540.7 + 24 \cdot 156.25 + \frac{25 \cdot 28}{53} \cdot (35 - 55.5)^2\right] = 459.6$$

Posterior: $\mathrm{NormalGamma}(45.8, 53, 459.6, 52)$

The Normal-Gamma Family

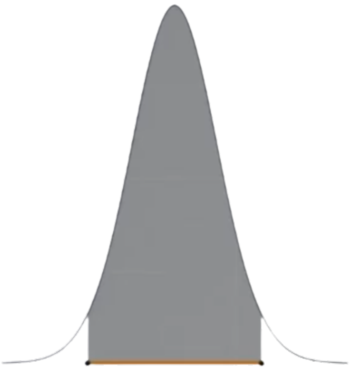
Example

Let's try to infer the total trihalomethane, or TTHM, in tap water from a city in North Carolina. Using prior information about TTHM from the city, as prior we shall use a NormalGamma distribution with a prior mean of 35 parts per billion based on a prior sample size of 25 and the estimate of the variance of 156.25 with degrees of freedom 24. In a new sample of size 28, the mean value of TTHM is 55.5 parts per billion with a sample variance of 540.7

Credible interval: (L, U) where $P(L \leq \mu \leq U \mid data) = 0.95$

$$L = t_{0.025}\sqrt{\frac{s_n^2}{n_n}} + m_n$$

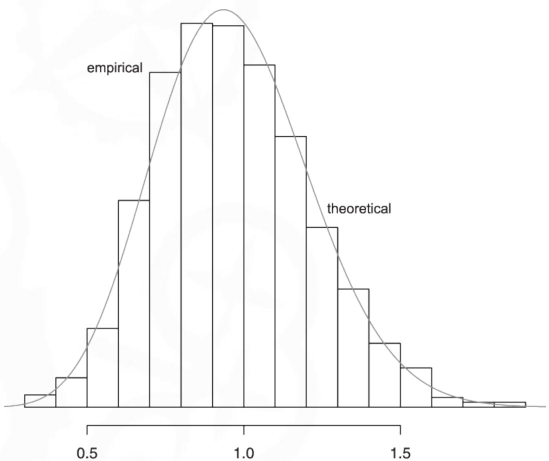
$$U = t_{0.975}\sqrt{\frac{s_n^2}{n_n}} + m_n$$



What happens if we are interested in distributions of other statistics apart from the mean?

Inference via Monte Carlo Sampling

- There may not be a closed form expression for other distributions.
- Suppose the posterior distribution is known, draw S random samples of ϕ :
$$\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(S)} \sim \text{Gamma}\left(\frac{\nu_n}{2}, s_n^2 \cdot \frac{\nu_n}{2}\right)$$
- As S gets large, the empirical distribution of the samples approximates the posterior distribution.
- From the empirical distribution, statistics of ϕ can be retrieved.



Predictions via Monte Carlo Sampling

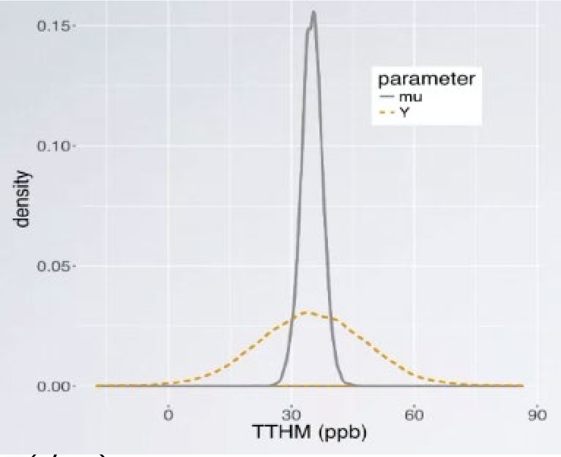
Example

For the examples of TTHM, we have been said to expect TTHM to be between 10-60 ppb. Based on the information, we may compute our priors:

$$m_0 = \frac{60+10}{2}$$

```
m_0 = (60+10)/2;  s2_0 = ((60-10)/4)^2;
n_0 = 25; v_0 = n_0 - 1
phi = rgamma(10000, v_0/2, s2_0*v_0/2)
sigma = 1/sqrt(phi)
mu =  rnorm(10000, mean=m_0, sd=sigma/(sqrt(n_0
y =   rnorm(10000, mu, sigma)
quantile(y, c(.025, .975))

##      2.5%     97.5%
##  8.522367 61.126357
```



$$\mu \sim N(m_0, \sigma^2/n_0)$$

$$Y \sim N(\mu, \sigma^2)$$

Predictions via Monte Carlo Sampling

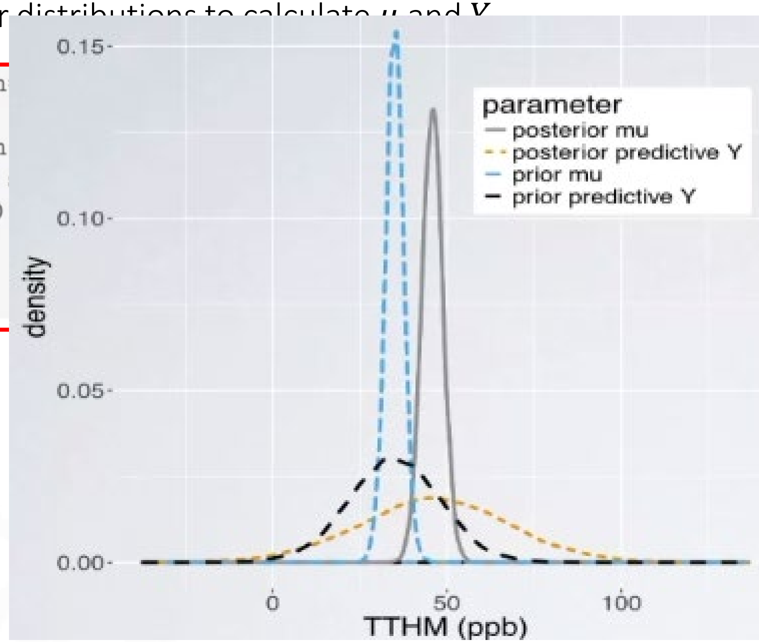
Example

For the examples of TTHM, we have been said to expect TTHM to be between 10-60 ppb. Based on the information, we may compute our priors:

We can also use the posterior distributions to calculate μ and Y

```
phi = rgamma(10000, v_n/2, s2_n
sigma = 1/sqrt(phi)
post_mu = rnorm(10000, mean=m_n
pred_y =  rnorm(10000,post_mu,
quantile(pred_y, c(.025, .975))

##     2.5%    97.5%
##  1.75404 89.30412
```



Predictive Inference (II)

Mixtures of Conjugate Priors

- In many situations, there may be reasonable prior information about the mean, but no reliable information about how many samples it represents:

$$\mu \mid \sigma^2, n_0 \sim N(m_0, \frac{\sigma^2}{n_0})$$

- This might be solved by using a hierarchical distribution to estimate the number of samples:

$$n_0 \mid \sigma^2 \sim \mathrm{Gamma}(1/2, r^2/2)$$

If r=1, this corresponds to a expected prior sample size of 1.

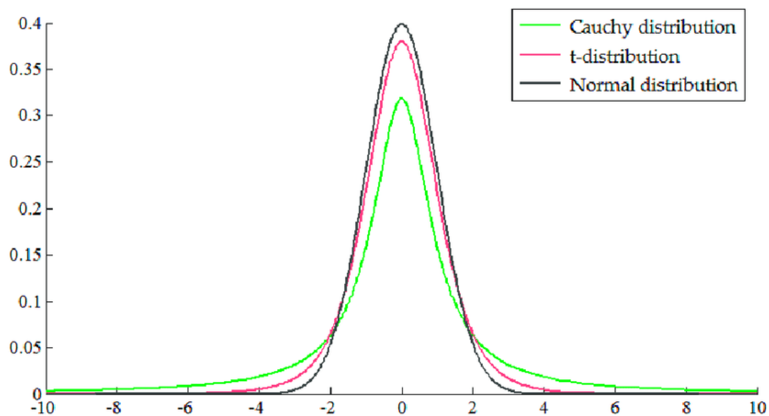
- It can be demonstrated via integration that, using this hierarchical distributions, distribution of μ is a **Cauchy distribution**:

$$\mu \mid \sigma^2 \sim \mathrm{C}(m_0, \sigma^2 \cdot r^2)$$

Mixtures of Conjugate Priors

- **The Cauchy distribution** is a special case of the student t with one DOF

$$p(\mu \mid \sigma) = \frac{1}{\pi \cdot \sigma \cdot r} \cdot \left(1 + \frac{(\mu - m_0)^2}{\sigma^2 \cdot r^2}\right)^{-1}$$



Predictive Inference (II)
Mixtures of Conjugate Priors

Posterior inference

- There is no nice closed form expression for posterior distribution of μ
- Conditional distributions of individual parameters given the others have closed form expressions.
- Use Markov Chain Monte Carlo alternating on each individual conditional distribution to converge on the posterior values for the parameters.

```
# initialize MCMC
sigma2[1] = 1; n_0[1]=1; mu[1]=m_0

#draw from full conditional distributions
for (i in 2:S) {
  mu[i]     = p_mu(sigma2[i-1], n_0[i-1],  m_0, r, data)
  sigma2[i] = p_sigma2(mu[i], n_0[i-1],    m_0, r, data)
  n_0[i]    = p_n_0(mu[i], sigma2[i],      m_0, r, data)
}
```

Predictive Inference (II)

Reference Priors

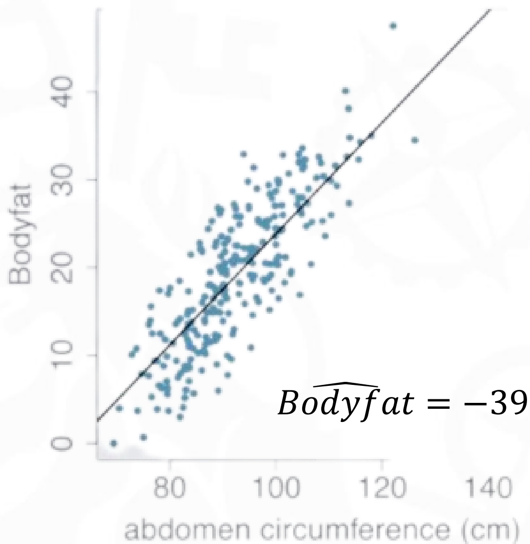
- A prior may have two roles: incorpórate existing or expert knowledge (**informative prior**) or represent minimal prior information due to unreliable information (**uninformative prior**).
- Uninformative prior are also called **objective priors**, which aim to minimize the influence of the prior on the posterior distribution.
- **Reference priors** are objective priors that **maximize the divergence between the prior and the posterior** distribution, allowing the data to play a more significant role in shaping the posterior distribution.
- Reference priors are applied when:
    - There is a lack of prior information or expert opinion
    - Trying to achieve consensus among different experts with conflicting opinions
    - Conducting sensitivity analyses to assess the impact of prior choice on the posterior

Bayesian Regression

Simple linear regression

Example

Data is available on the percent body fat and the abdominal circumference for 252 men. Picture shows the data with a fitted ordinary linear regression model.



Fitted values $\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i$

Residuals $\hat{\varepsilon}_i = y_i - \hat{y}_i$

σ^2 $MSE = \sum_{i=1}^{n} \frac{\hat{\varepsilon}_i^2}{n-2}$

$$\widehat{Bodyfat} = -39.28 + 0.63 \cdot Abdomen$$

Simple linear regression

Model and prior

Model follows the next structure:

$$Y_i = \alpha + \beta \cdot x_i + \varepsilon_i$$
$$\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

With marginal conditional distributions:

$$\text{cov}(\alpha, \beta \mid \sigma^2) = \sigma^2 \cdot S_{\alpha,\beta} \begin{cases} \alpha \mid \sigma^2 \sim N(a_0, \sigma^2 \cdot S_\alpha) \\ \beta \mid \sigma^2 \sim N(b_0, \sigma^2 \cdot S_\beta) \end{cases}$$
$$\frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{v_0}{2}, \frac{v_0 \cdot \sigma_0^2}{2}\right)$$

Reference priors:

$$uniform\ with\ p(\alpha, \beta, \sigma^2) \propto 1/\sigma^2$$

Simple linear regression

Posterior

Reference posterior:

$$\beta \mid y_1, \dots, y_n \sim t_{n-p-1}(\hat{\beta}, sd(\beta)^2)$$

$$\alpha \mid y_1, \dots, y_n \sim t_{n-p-1}(\hat{\alpha}, sd(\alpha)^2)$$

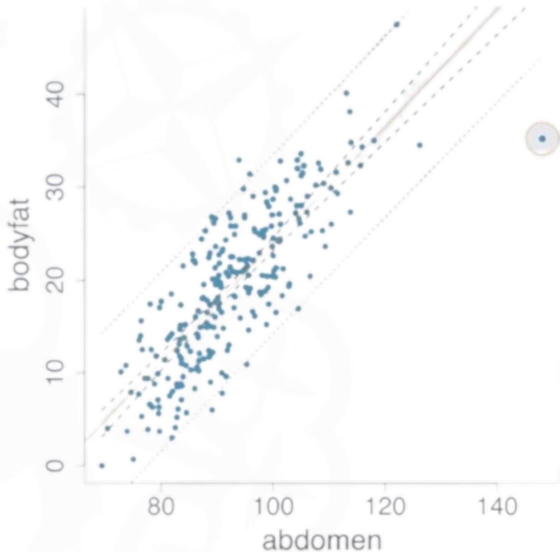
$$\alpha + \beta \cdot x_i \mid y_1, \dots, y_n \sim t_{n-p-1}(\hat{\alpha} + \hat{\beta} \cdot x_i, s_{y_i}^2)$$

$$s_{y_i}^2 = s_{Y|X}^2 \cdot \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$$

	Post. Mean	Post. Sd	2.5%	97.5%
Intercept	-39.28	2.66	-44.52	-34.04
Abdomen	0.63	0.03	0.58	0.69

Simple linear regression

	Post. Mean	Post. Sd	2.5%	97.5%
Intercept	-39.28	2.66	-44.52	-34.04
Abdomen	0.63	0.03	0.58	0.69



Bayesian Regression

Model selection critera

Bayesian Information Criterion (BIC)

$$-2 \cdot \log(likelihood) + \log(n) \cdot -2 \cdot \log(likelihood) + \log(n) \cdot parameters$$

- Smaller is better
- It involves a trade-off goodness of fit with model complexity.
- In regression:

$$-2 \cdot \log(1 - R^2) + \log(n) \cdot \#parameters$$

- Model uncertainty might be calculated using the posterior probabilities of the real data.

Alberto Aguilera 23, E-28015 Madrid - Tel: +34 91 542 2800 - http://www.iit.comillas.edu