$Bayesian Analysis

This course is hosted by Dr. Anita Wang, here is the course revision MD. File

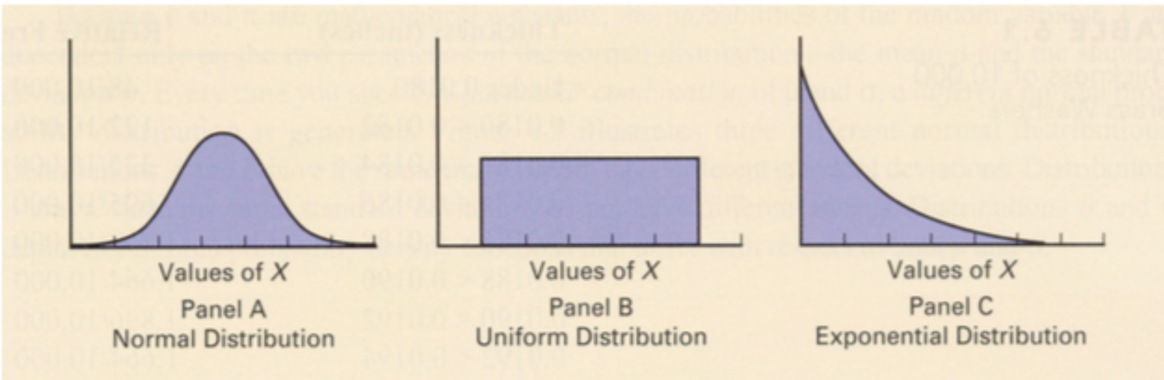
Random Variables

There are two kinds of random Variables: (we classify them with different distributions)

1. Continuous random variables:
    1. Uniform
    2. Univariabte Normal
    3. Gamma
    4. Beta
    5. Exponential
2. Discrete random variables

Continuous Random Variable

- A random variable is called **continuous** if the distribution function is continuous is differenctiable everywhere with the possible exception of a countable number of values.



Properties and Moments

The properties of those continuous distribution:

- $f(x) \geq 0$ for all x
- $\int xf(x) \ d(x) = 1$

The Expectation:

- $E(X) = \int x(fx) \ dx$

Variance:

- $Var(X) = \int x^2 f(x) dx - [E(X)]^2 = E(X^2)$

Uniform Distribution

The uniform distribution is used to represent a varaible that is **known to lie in an interval and equally likely** to be found anywhere in the interval .

- $X \sim U(\alpha, \beta)$, boundaries, α, β with $\beta < \alpha$
- $f(X) = \frac{1}{\beta-\alpha}$
- $E(X) = \int_{\alpha}^{\beta} \frac{1}{\beta-\alpha} x \ dx = \frac{\alpha-\beta}{2}$
- $Var(X) = \int_{\alpha}^{\beta} \frac{1}{\beta-\alpha} x^2 \ dx - (\frac{\beta+\alpha}{2})^2 = \frac{(\beta-\alpha)^2}{12}$

Univariate Normal Distribution

The Normal, or Gaussian, distribution is ubiquitous(adj. 普遍的) in statistics. Sample averages are approximately normally distributed by the **central limit theorem**.

- $X \sim N(\mu, \sigma^2)$, location μ, scale $\sigma > 0$
- $f(X) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$
- $E(X) = \int_{\alpha}^{\beta} x \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \ dx = \mu$
- $Var(X) = \int_{\alpha}^{\beta} x \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \ dx - \mu^2 = \sigma^2$

Properties:

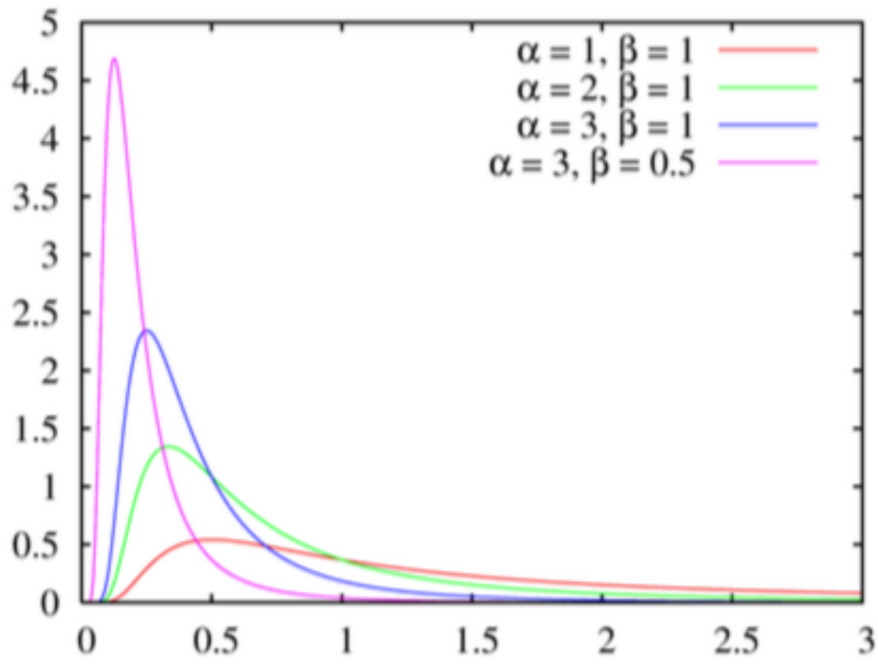
- The sum of two independent normal random varaibles is normally distributed. If X_1 and X_2 are independent with $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma^2 + \sigma^2)$
- If $X_1|X_2 \sim N(X_2\sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. Then $X_1 \sim N(\mu_2, \sigma_1^2 + \sigma_2^2)$

```
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt

X_nor=np.arange(-5,5,0.1)
mu=0
sigma=1

y=stats.norm.pdf(X_nor,mu,sigma)
plt.plot(X_nor,y)
```

Gamma Distribution

$X \sim Gamma(\alpha, \beta)$, shape $\alpha > 0$, inverse scale $\beta > 0$

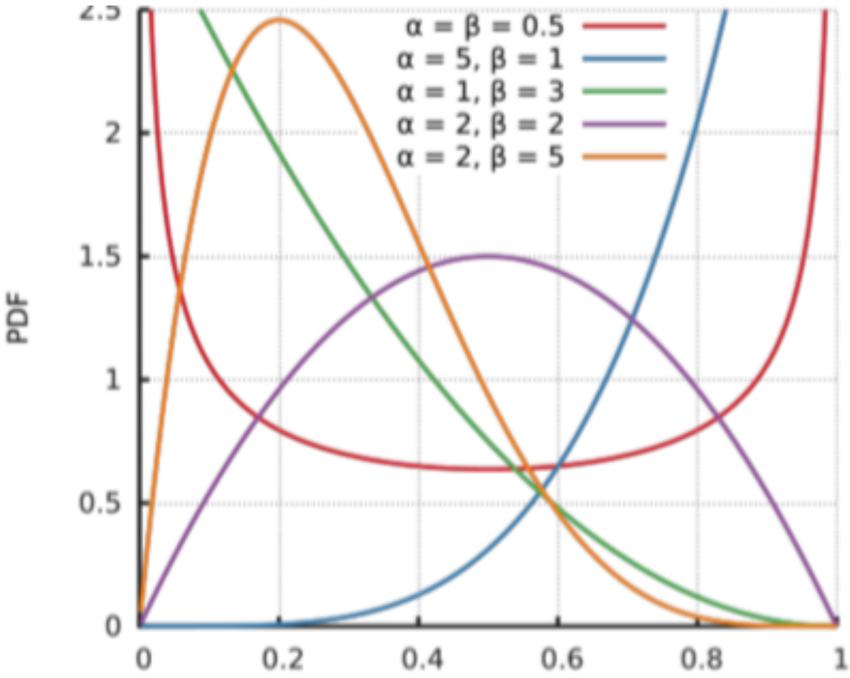
伽玛分佈中的参数 α，稱為形狀參數，β 稱為尺度參數



- $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},\ x > 0$
- $E(X) = \int_{-\infty}^{+\infty} x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\ dx = \frac{\alpha}{\beta}$
- $Var(X) = int_{-\infty}^{+\infty} x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\ dx - (\frac{\alpha}{\beta})^2 = \frac{\alpha}{\beta^2}$
- Property:
  - If X_1 and X_2 are independent with Gamma(α_1, β) and Gamma (α_2, β) distributions, then $X_1 + X_2 \sim \text{ Gamma } (\alpha_1 + \alpha_2, \beta)$

Beta Distribution

$X \sim Beta(\alpha, \beta)$, shape $\alpha > 0$, shape $\beta > 0$

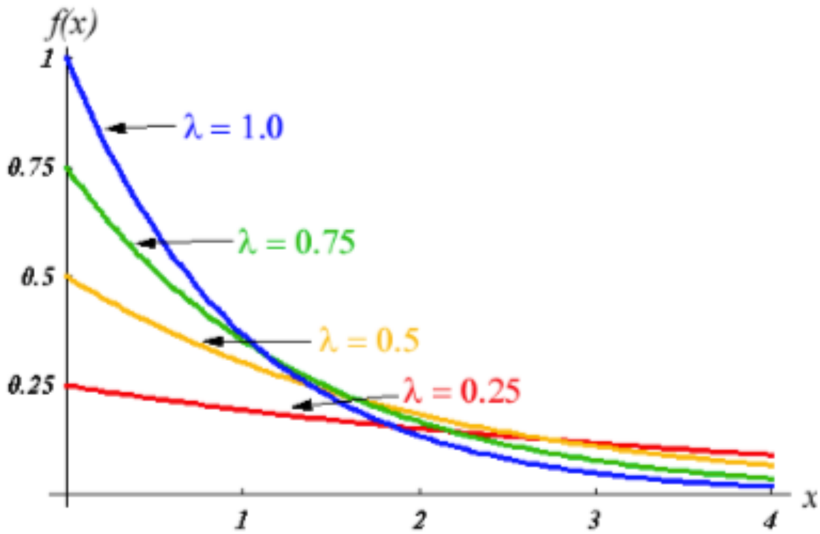


- $f(X) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, x \in [0,1], B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
- $E(X) = \int x \frac{x^{\alpha - 1} (1 - x)^{\beta - 1}}{B(\alpha, \beta)} \ dx = \frac{\alpha}{\alpha + \beta}$
- $Var(X) = \int x^2 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \ d(x) - (\frac{\alpha}{\alpha+\beta}) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Exponential Distribution

The exponential didstribution is the distribution of waiting times for the nex event in a Poisson process and is a special case of the gamma distribution with $\alpha = 1$.

- $X \sim Expon(\lambda)$, inverse scale $\lambda > 0$



- $f(X) = \lambda e^{\lambda x}, x > 0$

- $E(X) = \int_0^{+\infty} x\lambda e^{\lambda x}\ dx = \frac{1}{\lambda}$
- $Var(X) = \int_0^{+\infty} x^2\lambda e^{\lambda x}\ dx - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$

Discrete Random Variable

- The support of a random varaible is the set of numbers that are possible values of the random variable.
- A random variable is called discrete if the support contatins at most a **countable number values**

- Properties of probability function
  - For any values x of the random variable, $p(x) \geq 0$
  - The probabilities of all the events in the sample sapce must sum to 1, that is, $\sum\limits_{all} p(x) = 1$
- Expectation $E(X) = \sum\limits_{all x} x \times p(x)$
- Variance $Var(X) = \sum\limits_{all x} (x - E(x))^2 p(x)$

Example

- Let Y be a discrete random variable with probability function given in the following tatble

y_i	$f(y_i)$
0	0.20
1	0.15
2	0.25
3	0.35
4	0.05

  - Find $E(y)$
  - Find $Var(Y)$

Answer:

- $E(Y) = 0 \times 0.19 + 1 \times 0.15 + \cdots = 1.90$
- $Var(Y) = (0 - 1.90)^2 \times 0.20 + (1 - 1.90)^2 \times 0.20 \cdots = 1.49$

Poission Distribution

The Poisson distribution is commonly used to represent count data, such as the number of arrivals in a fixed time period.

- $X \sim Poisson(\lambda)$
- $p(x) = \frac{1}{x!}\lambda^x exp(-\lambda),\ x = 0, 1, 2, \ldots$
- Expectation: $E(X) = \sum_{k \geq 0} k \frac{1}{k!} \lambda^k e^{-\lambda} = \lambda^k e^{\lambda} \sum_{k \geq 1} \frac{1}{k-1!} \lambda^{k-1} = \lambda e^{-\lambda} e^{\lambda} = \lambda$
- Similarly, the variance is $Var(X) = E(X)^2 - [E(X)]^2 = \lambda$

Property

- If X_1 and X_2 are independent with Poisson (λ_1) and Poisson (λ_2) distribution, then $X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2)$

Binomial Distribution

The binomial distribution is commonly used to represent the number of **successes** in a sequence of n independent and identically distributed Bernoulli trials, with probability of success p in each trial.

- $X \sim Bin(n, p)$
- $p(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, 2, \ldots, n$
- Expectation: $E(X) = np$
- Variance: $Var(X) = np(1-p)$

Bayes' Rule and Inference

The **joint** probability mass or density function can be written as a product of the prior distirbution $p(\theta)$ and **the sampling distribution** $p(y|\theta)$

$$\mathbf{p(\theta, y) = p(\theta) p(y|\theta)}$$

Bayes' Rule:

The **posterior density**: $\mathbf{p(\theta_i|y) = \frac{p(\theta_i, y)}{p(y)} = \frac{p(\theta_i)p(y|\theta_i)}{p(y)}}$

Where $p(y) = \sum_{all\ \theta} p(\theta)p(y|\theta)$ over all possible values of θ (or $p(y) = \int_{all\ \theta} p(\theta)p(y|\theta)d\theta$) in the case of continuous θ).

*先验概率：**我们在未知条件下对事件发生可能性猜测的概率表示.。***

后验概率: 事情已经发生，要求这件事情发生的原因是由某个因素引起的可能性的大小

Example: College Students Sleeping

Prarameter p: the proportion of American college students who sleep at least eight hours.

A sample of 27 students is taken. In this group, 11 record that they had at least eight hours of sleep the previous night:

$$prior \ probability =\left\{ \begin{aligned} Pr(p = 0.2) & = & 0.6 \\ Pr(p=0.4) & = & 0.3 \\ Pr(p = 0.7) & = & 0.1 \end{aligned} \right.$$

The posterior probability:

$$Pr(p = 0.2|y) = \frac{Pr(p = 0.2)Pr(y|p = 0.2)}{\sum Pr(p)\Pr(y|p)}$$
$$= \frac{0.6 * \binom{27}{11}0.2^{11}0.8^{16}}{0.6\binom{27}{11}0.2^{11}0.8^{16} + 0.3\binom{27}{11}0.4^{11}0.6^{16} + 0.1\binom{27}{11}0.7^{11}0.3^{16}} = 0.089$$

$$Pr(p = 0.4|y) = \frac{Pr(p = 0.4)Pr(y|p = 0.4)}{\sum Pr(p)\Pr(y|p)}$$
$$= \frac{0.3 * \binom{27}{11}0.4^{11}0.6^{16}}{0.6\binom{27}{11}0.2^{11}0.8^{16} + 0.3\binom{27}{11}0.4^{11}0.6^{16} + 0.1\binom{27}{11}0.7^{11}0.3^{16}} = 0.909$$

$$Pr(p = 0.7|y) = \frac{Pr(p = 0.7)Pr(y|p = 0.7)}{\sum Pr(p)\Pr(y|p)}$$
$$= \frac{0.1 * \binom{27}{11}0.7^{11}0.3^{16}}{0.6\binom{27}{11}0.2^{11}0.8^{16} + 0.3\binom{27}{11}0.4^{11}0.6^{16} + 0.1\binom{27}{11}0.7^{11}0.3^{16}} = 0.002$$

Bayesian Thinking

- Parameter θ is unknown and to be estimated
- Previously, we used sample data information to estimate θ (Fro example, sample proportion $\hat{p}$ to estimate population proportion p)
- Bayesian thinking:
    - Prior information of the parameter: the subject prior opinion ofthe distribution of the parameter
    - Sample data information
    - Posterior distribution: combine the information in the data with the prior dstirbution

Satistical Inference

Two main approaches

- Frequentist
    - Model parameters are fixed unknown quantites.
        - Estimation - MLE, MME
        - Confidence intervals
        - Significance testing - p-values
        - Hypothesis testing -Reject / Don't Reject H_0
- Bayesian

- Model parameter are random varaibles. Inference is based on $P(\theta|Data)$, the posteriro distribtuion given the data
  - Estimation - Posterior menas, modes
  - Credible intervals/sets
  - Posteriro probabilities

An equivalent form omits the factor $p(y)$, which does not depend on θ and ,with fixed y, can thus be considered a constant, yielding the **unnormalized posterior density**:

$$\underbrace{\mathbf{p(\theta|y)}}_{\mathbf{posteriordensity}} \propto \underbrace{\mathbf{p(\theta)}}_{\mathbf{prior}} \overbrace{\mathbf{p(y|\theta)}}^{\mathbf{Likelihood}}$$

The second term in this expression, $p(y|\theta)$, is taken here as a function of θ, not of y.

Prediction

Prior predictive distribution (Also called marginal distribution of y)

$$p(y) = \int p(y,\theta)\ d\theta = \int p(\theta)p(y|\theta)d\theta$$

prior because it is not conditional on a previous observation of the process, and predictive because it is the distribution ofr a quantity that is observable.

Posterior predictive distribution

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y},\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta,y)p(\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \end{aligned}$$

Once the data y have been observed, the unknown observable $\tilde{y}$ can be predicted. For example, $y = \{y_1, y_2, \ldots, y_n\}$ may be the vector of recorded weights of an object weighed n times on a scale, $\theta = (\mu, \sigma^2)$ is the prior, and $\tilde{y}$ may be the yet to be recorded weight of the object in a planned new weighing.

Simulated the Posterior Predictive Distribution

- Assuming that you can simulate from the posterior distribution of the parameter, which is usually feasible.
- To simulate the posterior predictive distribution involves two steps:
  - Simmulate θ_i from $\theta|y; i = 1, 2, \ldots, m$
  - Simulate $\tilde{y}_i$ from $\tilde{y}|\theta_i, y)$

- The pairs $(\theta_i, \tilde{y_i})$ are draws from the joint distribution $Pr(\theta, \tilde{y}|y)$. Therefore, the $\color{red}{\backslash\text{thilde}}y_i$ are draws from $\tilde{y}|y$

Single Parameter Model

- Single parameter model is statistical models where only a single scalar parameter is to be estimated; that is, the estimated θ is **one-dimensional**

- In this chapter:
    - Binomial
    - Normal
    - Poisson
    - Exponential

Binomial

- In this simple binomial model, the aim is to estimated an **unknown population proportion from the results of a sequence of 'Bernoulli trials'; that is , data $y_1, \ldots, y_n$.

- Because of the exchangeability, the data can be summarized by the total number of succeses in the n trails, which we denote here by y.

- The binomial sampling distribution is $p(y|\theta) = Bin(y|n, \theta) = \binom{n}{y} p^y (1-p)^{n-y}$, where on the left side we suppress the dependence on n becasue ti is regarded as part of the experimental design that is considered fixed.

- **Example**
    - We consider the estimation of the sex ratio within a population of human births. The currently accepted value of the proportion of female births in large European-race populations is 0.485.
    - Let y be the number of girls in n recorded births. We are assuming that the n births are conditionally independent given θ, with the probaibility of a female birth eqaul to θ for all cases
    - For simplicity, we assume that the prior distirbution for θ is **uniform** on the interval $[0, 1]$.
    - The posterior density: $p(y|\theta) = Bin(y|n, \theta) = \binom{n}{y} p^y (1-p)^{n-y}$

Different prior densities

We consider a parametric family of prior distributions that includes the uniform as a special case and construct a family of prior densities that lead to simple posterior densities.

- $\theta \sim Beta(\alpha, \beta)$: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$
- The posterior density:

$$\begin{aligned} p(\theta|y) &\propto \theta^y (1-\theta)^{n-y} \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \\ &= Beta(\theta|\alpha+y, \beta+n-y) \end{aligned}$$

Conjugate prior

The property that the posterior distribution foloows the same parametric form as the prior distribution is called **conjugacy**; the beta prior distribution is a **conjugate family** for the binomial likelihood.

- If F is a class of sampling distributions $p(y|\theta)$, and P is a class of prior distributions for θ, then the calss P is conjugate for F if $p(\theta|y) \in P$ for all $p(y_i|\theta) \in F$ and $p(y_i) \in P$
- This definition is formally vage since if we choose P as the class of all distributions, then P is always conjugate no matter what class of sampling distribution is used.

Normal mean with known variance: a single observation

Consider a single scalar observation y from a normal distribution parameterized by a mean θ and variance σ^2, where for this initial develpment we assume that σ^2 is known.

$\mathbf{p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma}} exp\{-\frac{(y-\theta)^2}{2\sigma^2}\}}$

The conjugate prior θ $N(\mu_0, \tau_0^2)$

$p(\theta) =\propto (-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2)$

hyperparameters 、mu_0 and τ^2

$$
\begin{aligned}
p(\theta|y) &\propto p(\theta) \times p(y|\theta) \\
&= (-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2) exp\{-\frac{(y-\theta)^2}{2\sigma^2}\}
\end{aligned}
$$

$\sigma|y \sim N(\mu_1, \tau_1^2)$ Where $\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{y}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}$ and $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$

the posterior **precision** equals the prior precision plus the data precision

Mulitiple Observations

a sample of independent and identically distributed observations $y = (y_1, y_2, \ldots, y_n)$ is available

- Posterior density:

$$
\begin{aligned}
p(\theta|y) &\propto p(\theta)p(y|\theta) \\
&= p(\theta) \prod_{i=1}^{n} exp(-\frac{1}{2\sigma^2}(y_i - \theta)^2) \\
&\propto exp(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2) \prod_{i=1}^{n} exp(-\frac{1}{2\sigma^2}(y_i - theta)^2) \\
&\propto exp(-\frac{1}{2}(\frac{1}{\tau_0^2}(\theta - \mu_0)^2) + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2)
\end{aligned}
$$

Posterior distribution

The posterior distribution is also a normal distribution:

$\sigma|y \sim N(\mu_n, \tau_n^2)$ Where $\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$ and $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$

Normal distribution with known mean but unknown variance

For $p(y|\theta, \sigma^2) = N(y | \theta, \sigma^2)$, with θ known and σ^2 unknown, the likelihood for a vector y of n i.i.d. observations is:

$$
\begin{aligned}
P(y|\sigma^2) &\propto \sigma^{-n} exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2) \\
&= (\sigma^2)^{-\frac{n}{2}} exp(-\frac{n}{2\sigma^2}v)
\end{aligned}
$$

The sufficient statistics is $v = \frac{1}{n}\sum_{i=1}^{n}(y_i-\theta)^2$

正态分布来说，充分统计量，一个均数，一个是方差。

Poisson distribution

- Observations: $y = (y_1, y_2, \ldots, y_n)$
- Likelihood:

  $$\mathbf{p(y|\theta) = \prod_{i=1}^{n}\frac{\theta^{y_i}e^{-\theta}}{y_i} \propto \theta^{n\bar{y}}e^{-n\theta}}$$

- Prior density: Gamma(α, β)
- Posterior density:

  $$\mathbf{p(\theta|y) \propto e^{-(n+\beta)\theta}\theta^{n\bar{y}+\alpha-1}}$$

  $$\mathbf{\theta|y \sim Gamma(n\bar{y}+\alpha, n+\beta)}$$

Exponential Distribution

- Observations: $y = (y_1, y_2, \ldots, y_n)$
- Likelihood:

  $$\mathbf{p(y|\theta) = \theta^{n}exp[-n\bar{y}\theta]}$$

- Prior density: Gamma(α, β)

  $$\mathbf{p(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta}}$$

- Posterior density

  $$\mathbf{p(\theta|y) \propto \theta^{\alpha+n-1}exp(-(n\bar{y}+\beta)\theta)}$$

  $$\mathbf{\theta|y \sim Gamma(\alpha+n), n\bar{y}+\beta)}$$

The sampling distribution when viewed as the likelihood of θ, for fixed y, is proportional to a Gamma($n+1, ny$) density. Thus the Gamma(α, β)prior distirbution for θ can be viewed as $\alpha-1$ exponential observations with total waiting time β

Prior

Conjugate and Non-Conjugate

A prior is **conjugate** if the posterior is a member of the same parametric family. The advantage is that the posterior is available in closed form (very convenient for MCMC)

For many likelihoods/parameters, there is no known conjugate prior

A silly Example:

Say $Y \sim Poisson(\lambda)$ and $\lambda \sim Beta(a, b)$

Infromative and Uninformative priors

Conjugacy is only about the family of distirbutions that the prior belongs to. However, we can utilize outside knowledge to make a prior (conjugate or non-conjugate more) **informative**.

- Potential sources include literature reviews, pilot studies, and expert opinions.
- **Prior elicitation** is the process of converting expert information into a prior distribution
  - For example, the expert may not know what an inverse gamma pdf is , but you can choose a and b so that the distribution reflects the beliefs of the expert.
- Any time onformative priors are used you shoud conduct a **sensitivity analysis**, that is, evaluate how the posterior differs for several priors.

Subjective and Objective

- Subjective decisions include picking the likelihood, treatment of outliers, transformation, $\ldots$ and prior specification
- An **objective analysis** is one that requires no subjective decisions by the analyst

Jeffreys' Priors

- Jeffreys' principle leads to defining hte **noninformative prior density**

  $p(\theta) = [J(\theta)]^{\frac{1}{2}}$

  Where $J(\theta)$ is the Fisher information for θ

  $\mathbf{J}(\theta) = \mathbf{E}[(\frac{\mathbf{d\ log\ p(y|\theta)}}{\mathbf{d}\theta})^\mathbf{2}|\theta] = -\mathbf{E}[(\frac{\mathbf{d^2\ log\ p(y|\theta)}}{\mathbf{d}\theta^\mathbf{2}})|\theta]$

Multiparameter Model

Introduction

virtually every practical problem in satistics involves more thatn one unknown or unobservable quantity.

The ultimate aim of Bayesian analysis is to obtain the **marginal posterior distribution** of the particular parameters of interest.

We firs require the joint posterior distribution of all unknowns, and then we integrate this distribution over the unknown that to obtain the desired maginal distribution.

In many problems there is **no interst in making inferences about many of the unknown parameter**. Parameters of thsi kind are often called **nuisance parameters**.

Averaging over 'nuisance parameters'

- Suppose θ has two parts, each of which can be a vector, $\theta = (\theta_1, \theta_2)$
- supppose that we are only interested (at least for the moment) in inference for θ_1, so θ_2 may be considered a **'nuisance' parameter**
- For instance, in the simple example

  $y|\mu, sigma^2 \sim N(\mu, \sigma^2)$

In which bouth $\mu(=' \ 01')$ and $\sigma_2(=' \ \theta_2')$ are unknown, interst commonly centers on μ.

joint Posterior Distribution

- We seek the conditional distirbution of the parameter of interst given the observed data; in this case, $p(\theta_1|y)$.
- Joint posterior density:
  - $p(\theta_1, \theta_2|y) \propto \int p(\theta_1, \theta_2|y) d\theta)2$
- Marginal posterior density:
  - $p(\theta_1|y) = \int p(\theta_1, \theta_2|y) d\theta_2$
- Alternatively,
  - $p(\theta_1|y) = \int p(\theta_1|\theta_2, y) p(\theta_2|y) d\theta_2$

Marginal Posterior Density

- $p(\theta_1|y)$ can be regarded as a mixture of the conditional posterior distirbutions given the nuisance parameter, θ_2, where $p(\theta_2|y)$ is a weighting function for the different possible values of θ_2
- The weights depend on the posterior density of θ_2 and thus on a combination of evidence from data and prior model
- $p(\theta_1|y)$ can be computed by marginal and comditional simulation, first drawing θ_2 from its maringal posterior distribution and then θ_1 from its conditional posterior distribution, given the dran values of θ_2

Univariate Normal with a Noninformative Prior

Consider a vecotr y of n independent observations from a univariate normal distribution, $N(\mu, \sigma^2)$

Assuming prior independence of location and scale parameters, if uniform on $(\mu, log\ \sigma)$or, $p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$

The joint posterior distribution: $p(\mu, \sigma^2|y) \propto \sigma^{-n-2} exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2)$