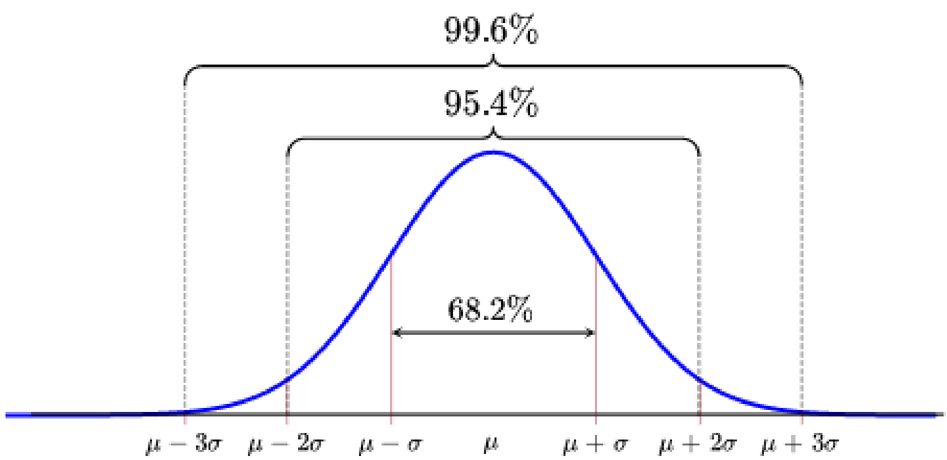
Stat 136 (Bayesian Statistics)

Lesson 2.1: Bayesian Inference for the Mean of a Normal Distribution

Introduction: The Normal Distribution

We have learned that many random variables follow the normal distribution or at least can be approximated by a normal distribution. The normal distribution is a symmetric, bell shaped distribution and is parameterized by mean μ and variance σ^2. If $Y \sim N(\mu, \sigma^2)$, then

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$



Bayes inference for the normal mean using continuous prior

We know that Bayes' theorem can be summarized as posterior proportional to prior times likelihood. Suppose we have a random sample $Y_1, Y_2, \cdots, Y_n$ from $N(\mu, \sigma^2)$, where σ^2 is known. It is more realistic to believe that all values of μ are possible, at least all those in an interval. This means we should use a continuous prior. Some possible prior distributions are:

- Jeffrey's uniform prior: a flat (non-informative) prior, also an improper prior, $f(\mu) = 1$
- Normal prior: a conjugate prior (Normal-Normal conjugate pair)

Jeffrey's uniform prior

The Jeffrey]s prior for the mean is given by $f(\mu) = 1$ a prior which gives each possible value of μ equal weight. In other words, it does not favor any value over any other value. This prior is not a proper prior distribution since it cannot integrate to 1. Nevertheless, this improper prior works out all right. Even though the prior is improper, the posterior will integrate to 1, so it is proper. The Jeffrey's prior for the mean of a normal distribution turns out to be the a flat prior.

Now, given a random sample $Y_1, Y_2, \cdots, Y_n$ from $N(\mu, \sigma^2)$, the likelihood function is given by

$$\begin{aligned} L(\mathbf{y}|\mu,\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}, \text{ where: } \mathbf{y} = [y_1, y_2, \cdots, y_n] \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n exp\left[-\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2\sigma^2}\right] \\ &\propto exp\left[-\frac{1}{2(\sigma^2/n)}(\overline{y}-\mu)^2\right] \end{aligned}$$

It can be shown that

$$\begin{aligned} \sum_{i=1}^{n}(y_i-\mu)^2 &= \sum_{i=1}^{n}(y_i-\overline{y})^2 + \sum_{i=1}^{n}(\overline{y}-\mu)^2 \\ &\propto n(\overline{y}-\mu)^2 \end{aligned}$$

Thus, the likelihood of a random sample of $Y_i \sim N(\mu, \sigma^2)$ can be expressed as the likelihood of $\overline{y}$. But, we knew that $\overline{y} \sim N(\mu, \sigma^2/n)$.

Now, using the Bayes Theorem the posterior distribution is identical to the likelihood and is given by

$$f(\mu|\overline{y}) \propto exp\left[-\frac{1}{2(\sigma^2/n)}(\mu-\overline{y})^2\right]$$

Normal prior density for the mean of normal distribution

Suppose we consider a normal density as the prior for μ, say $N(\theta, \tau^2)$. That is,

$$f(\mu) \propto exp\left[-\frac{(\mu-\theta)^2}{2\tau^2}\right]$$

Hence, the posterior distribution of μ is

$$
\begin{aligned}
f(\mu|\bar{y}) &= f(\mu) \times L(\bar{y}|\mu, \sigma^2) \\
&\propto exp\left[-\frac{(\mu-\theta)^2}{2\tau^2}\right] \times exp\left[-\frac{(\bar{y}-\mu)^2}{2\sigma^2/n}\right] \\
&\vdots \\
&= exp\left[-\frac{1}{2\left(\frac{\tau^2\sigma^2}{\sigma^2+n\tau^2}\right)}\left(\mu - \frac{\sigma^2\theta + n\tau^2\bar{y}}{\sigma^2+n\tau^2}\right)^2\right]
\end{aligned}
$$

This means that μ is distributed as $N(\mu^*, \sigma^{2*})$, where:

$$
\begin{aligned}
\mu^* &= \frac{\sigma^2\theta + n\tau^2\bar{y}}{\sigma^2+n\tau^2} \\
&= \left(\frac{\sigma^2}{\sigma^2+n\tau^2}\right)\theta + \left(\frac{n\tau^2}{\sigma^2+n\tau^2}\right)\bar{y} \\
&= \left(\frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)\theta + \left(\frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)\bar{y}
\end{aligned}
$$

Thus, it is clear that the posterior mean μ^* is the weighted average of the prior mean μ_p and $\bar{y}$, where the weights are the proportions of the posterior precision.

Meanwhile, the posterior variance is informed by the prior variability τ^2 and variability in the data σ62. Further, as n increases, the posterior variance decreases. That is, the more and more data we have, our posterior certainty about μ increases and becomes more in sync with the data.

$$
\sigma^{2*} = \frac{\tau^2\sigma^2}{\sigma^2+n\tau^2} \implies \frac{1}{\sigma^{2*}} = \frac{n}{\sigma^2} + \frac{1}{\tau^2}
$$

NOYTE: When σ is unknown we estimate it using the sample standard deviation, s, and recalculate μ^* and σ^* with σ replaced by s.

Example

Arnie and Marie are going to estimate the mean length of tilapia in a fishpond. Previous studies in other fishponds have shown the length of tilapia to be normally distributed with known standard deviation of 2 cm. Hence, the likelihood function is

$$
y_i \sim N(\mu, 2^2) \implies \bar{y} \sim N(\mu, \frac{2^2}{12})
$$

Arnie decides his prior mean is 30 cm. He does not believe it is possible for a yearling rainbow to be less than 18 cm or greater than 34 cm. Thus, his prior standard deviation is approximately 4 cm. Thus, he will use a $N(30, 16)$ prior. On the other hand, Marie does not know anything about tilapia, so she decides to use the Jeffrey's prior.

- Arnie's prior: $f(\mu) \approx exp\left[-\frac{(\mu-30)^2}{2(16)}\right]$, (a normal prior)
- Marie's prior: $f(\mu) = 1$ (Jeffrey's flat, non-informative prior)

They take a random sample of 12 tilapia from the pond and calculated the sample mean to be $\bar{y} = 32cm$.

Consequently for Marie, since she used Jeffrey's flat prior, the posterior distribution is equal to the likelihood, that is, $f(\mu|\bar{y}) \sim N(\mu^*, \sigma^{2*})$, where

$$\mu^* = \bar{y} = 32$$

and

$$\sigma^{2*} = \frac{\sigma^2}{n} = \frac{2^2}{12} = \frac{1}{3}$$

While, for Arnie, the posterior distribution is also $N(\mu^*, \sigma^{2*})$, where the posterior mean and variance are, respectively,

$$\mu^* = \frac{\sigma^2\theta + n\tau^2\bar{y}}{\sigma^2 + n\tau^2} = \frac{2^2(30) + 12(16)(32)}{2^2 + 12(16)} \approx 31.96$$

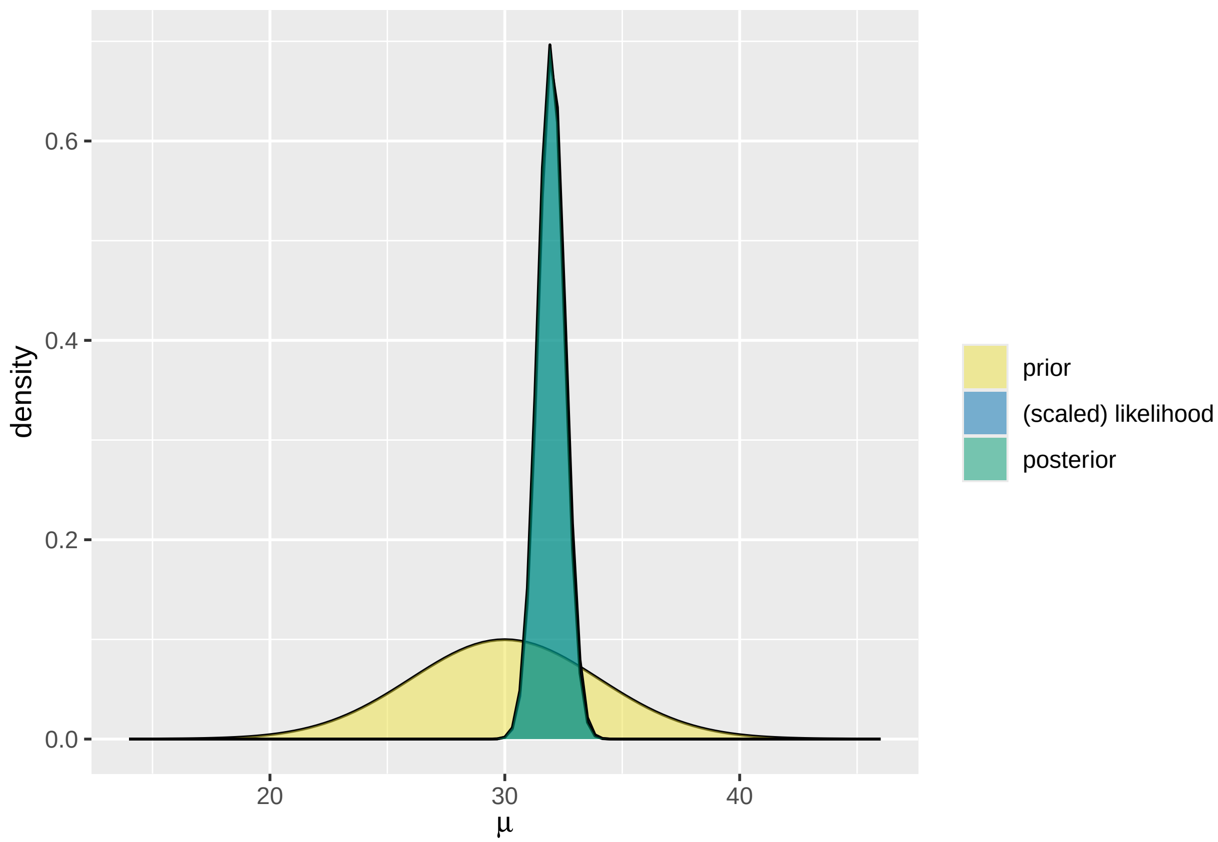
and

$$\sigma^{2*} = \frac{\sigma^2\sigma_p^2}{\sigma^2 + n\sigma_p^2} = \frac{2^2(16)}{2^2 + 12(16)} \approx 0.3265$$

Implementation in R

There are a few functions in the **bayesrules** package which can facilitate the above calculations. For example, the *plot_normal_normal()* and *summarize_normal_normal()* functions generate the distribution plot and summary statistics of the posterior, respectively

```
plot_normal_normal(mean = 30, sd = 4, sigma = 2,
                   y_bar = 32, n = 12)
```




```r
summarize_normal_normal(mean = 30, sd = 4, sigma = 2,
                        y_bar = 32, n = 12)
```

```
##       model     mean     mode        var        sd
## 1     prior 30.00000 30.00000 16.0000000 4.0000000
## 2 posterior 31.95918 31.95918  0.3265306 0.5714286
```

Choosing a normal prior

How does one, in practice, choose a Normal prior for μ that reflects prior beliefs about the location of this parameter? One indirect strategy for choosing for selecting values of the prior parameters μ_p and σ_p^2 is based on the specification of quantiles. On the basis of one's prior beliefs, one specifies two quantiles of the Normal density. Then, the Normal parameters are found by matching these two quantiles to a particular Normal curve. The matching is performed by the R function *normal.select()* in the **LearnBayes package**. Input the two quantiles by the *list()* statements, and the output is the mean and standard deviation of the Normal prior.

For example, suppose $P_{50} = 30$ and $P_{90} = 32$

```
normal.select(list(p=0.5,x=30),list(p=0.9,x=32))
```

```
## $mu
## [1] 30
##
## $sigma
## [1] 1.560608
```

In practice we perform several checks to see if this Normal prior makes sense. Several functions are available to help in this prior checking.

```
qnorm(0.25,30, 1.56)
```

```
## [1] 28.9478
```

The Bayes' theorem allows the updating of our prior information when data is already available. Without going thru the calculations demonstrated earlier, we can performed updating by the R function *normal_update()* in the **ProbBayes** package. One inputs two vectors: *prior* is a vector of the prior mean and standard deviation and *data* is a vector of the sample mean and standard error

The output is a vector of the posterior mean and posterior standard deviation

```
prior <- c(30, 4)
data <- c(32, 0.58)
normal_update(prior, data)
```

```
## [1] 31.9588159  0.5739972
```

Bayesian credible interval for the normal mean

Recall that the posterior distribution for μ is normal with mean μ^* and standard deviation σ^*. A $(1-\alpha)\times 100\%$ credible interval for μ is given by

$$\mu^* \pm z_{\frac{\alpha}{2}}\sigma^*$$

When σ is unknown we estimate it using the sample standard deviation (s) and the $(1-\alpha)\times 100\%$ credible interval for μ is given by

$$\mu^* \pm t_{\frac{\alpha}{2}}\sigma^*$$

Example: (continued)

For Marie, her 95% credible interval for μ is $32 \pm 1.96(0.5774) \implies (30.87, 33.13)$

- There is a 95% probability that $(30.87, 33.13)$ contains μ

For Arnie, his 95% credible interval for μ is $31.96 \pm 1.96(0.5714) \implies (30.84, 33.08)$

- There is a 95% probability that $(30.84, 33.08)$ contains μ

We can also construct credible instervals by simulating observation from the posterior distribution.

```
m <- rnorm(100000, 32, 0.5774)
a <- rnorm(100000, 31.96, 0.5714)
quantile(m,c(0.025,0.975)) #Marie
```

```
##     2.5%    97.5%
## 30.86556 33.12814
```

```
quantile(a,c(0.025,0.975)) #Arnie
```

```
##     2.5%    97.5%
## 30.83864 33.07763
```

Bayesian one-sided hypothesis test about μ

Recall that the posterior distribution $f(\mu|\bar{y})$ summarizes our entire belief about the parameter, after viewing the data. Suppose we want to test $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$.

Testing a one-sided hypothesis in Bayesian statistics is done by calculating the posterior probability of the null hypothesis, $P(H_0|\mu^*, \sigma^{2*})$.

$$\begin{aligned} P(H_0 : \mu \leq \mu_0|\mu^*, \sigma^{2*}) &= \int_{-\infty}^{\mu_0} f(\mu|\bar{y}) d\mu \\ &= P\left(Z \leq \frac{\mu^* - \mu_0}{\sigma^*}\right) \end{aligned}$$

If this probability is less than the α-level of significance, H_0 is rejected.

Example 3: (continued)

For example, Marie wants to test $H_0 : \mu \leq 31$ versus $H_1 : \mu > 31$

$$\begin{aligned} P(\mu \leq 31) &= P\left(Z \leq \frac{31 - 32}{0.5774}\right) \\ &= P(Z \leq -1.73) \\ &\approx 0.0418 \end{aligned}$$

Thus, we reject H_0 at $\alpha = 0.05$.

Alternatively, we can test the same hypothesis using simulated observations from the posterior distribution.

```
samples <- 100000
m <- rnorm(samples,32, 0.5774)
sum(m<=31)/samples
```

```
## [1] 0.04209
```

Let us test the same hypthesis based on the posterior distribution of Arnie.

$$
\begin{aligned}
P(\mu \leq 31) &= P\left(Z \leq \frac{31 - 31.96}{0.5714}\right) \\
&= P(Z \leq -1.68) \\
&\approx 0.0465
\end{aligned}
$$

Likewise, we reject H_0 at $\alpha = 0.05$. Using simulation data we arrive at the same conclusion.

```
samples <- 100000
m <- rnorm(samples,31.96, 0.5714)
sum(m<=31)/samples
```

```
## [1] 0.04689
```

Bayesian two-sided hypothesis test about μ

Suppose we want to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. In order to test two-sided hypothesis we construct the $(1 - \alpha) \times 100\%$ credible interval for μ.

- If μ_0 is contained in the credible interval, we do not reject H_0 and conclude that μ_0 still has credibility as a possible value for μ

For example, if we test $H_0 : \mu = 35$ against $H_1 : \mu \neq 35$ using Arnie's posterior distribution at $\alpha = 0.05$. Now, recall that the 95% credible for μ based on Arnie's posterior distribution is (30.84, 33.08) which does not include $\mu = 35$, hence, we reject H_0.

Bayesian prediction

Suppose we wanted to predict the length of a randomly selected tilapia. Again, we can use the posterior distribution to accomplish this task.

- Two-step procedure:
  - Sample a value of μ from its posterior distribution
  - Sample a new observation $\widetilde{Y}$ from the data model (i.e. a prediction)

```
mu_sim <- rnorm(100000, 31.96,0.5714)
y_sim <- rnorm(100000, mu_sim, 2 )
round(mean(y_sim),0)
```

```
## [1] 32
```

Therefore, based on this Bayesian model, the expected length of a randomly selected tilapia is 32 cm.