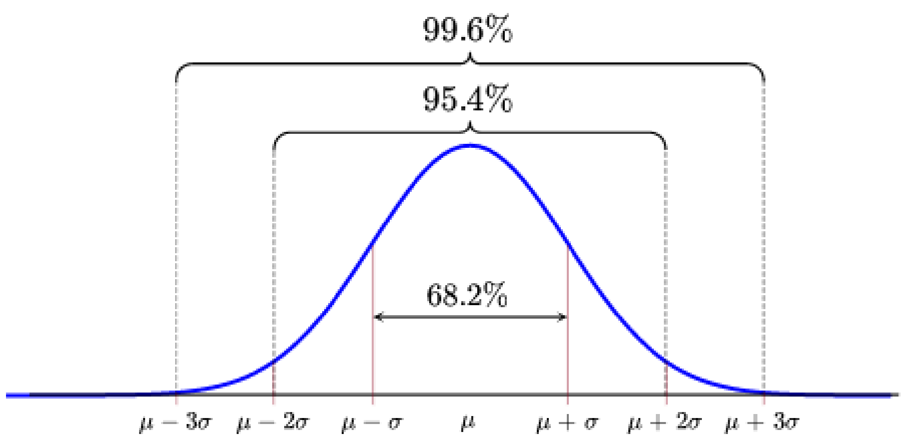
Stat 136 (Bayesian Statistics)

Lesson 3.1 Bayesian Inference for the Mean of a Normal Distribution

Introduction: The Normal Distribution

- Many random variables seem to follow the normal distribution, at least approximately
- The normal distribution is a symmetric, bell shaped distribution
- It has two parameters: 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 discrete prior

- Suppose there are m possible values of $\mu : \mu_1, \mu_2, \cdots, \mu_m$
- Choose a discrete prior probability distribution,$f(\mu)$, which summarizes our prior belief about the parameter
- Suppose a single observation $Y \sim N(\mu, \sigma^2)$, with σ^2 known

- For a single observation, the likelihood is given by

$$
\begin{aligned}
L(y|\mu) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\
&\propto e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\
&= exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]
\end{aligned}
$$

- For a random sample of observation $y_1, y_2, \cdots, y_n$, the joint likelihood is given by

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

Example 1:

Suppose $Y \sim N(\mu, 1)$. Suppose further that there are only five possible values for μ. These are 2.0, 2.5, 3.0, 3.5, and 4, and each is equally likely to occur. Let $Y = 3.2$.

```r
mu <-  c(2.0, 2.5, 3.0, 3.5, 4)
pr <- c(1/5, 1/5, 1/5, 1/5, 1/5)
lik <-  round(exp(-(3.2-mu)^2/2),4)
PrxL <- round(pr*lik,4)
post <- round(PrxL/sum(PrxL),4)
knitr::kable(cbind(mu, pr, lik, PrxL, post),
             col.names = c("mu",
                           "Prior",
                           "Likelihood",
                           "Prior x Likelihood",
                           "Posterior"))
```

mu	Prior	Likelihood	Prior x Likelihood	Posterior
2.0	0.2	0.4868	0.0974	0.1239
2.5	0.2	0.7827	0.1565	0.1990

mu	Prior	Likelihood	Prior x Likelihood	Posterior
3.0	0.2	0.9802	0.1960	0.2493
3.5	0.2	0.9560	0.1912	0.2432
4.0	0.2	0.7261	0.1452	0.1847

Example 2:

Suppose we take a random sample of 4 observations from a normal distribution having mean μ and variance $\sigma^2 = 1$. Let the observations be 3.2, 2.2, 3.6, and 4.1. Consider the same values of μ and discrete prior probabilities in *Example 1*.

```
yobs <- c(3.2, 2.2, 3.6, 4.1)
n <- length(yobs)
lik2 <-  round(exp(-n*(mean(yobs)-mu)^2/2),4)
PrxL2 <- round(pr*lik2,4)
post2 <- round(PrxL2/sum(PrxL2),4)
table2 <- as.data.frame(cbind(mu, pr, lik2, PrxL2, post2))
knitr::kable(table2,
             col.names = c("mu",
                           "Prior",
                           "Likelihood",
                           "Prior x Likelihood",
                           "Posterior"))
```

mu	Prior	Likelihood	Prior x Likelihood	Posterior
2.0	0.2	0.0387	0.0077	0.0157
2.5	0.2	0.3008	0.0602	0.1228
3.0	0.2	0.8596	0.1719	0.3505
3.5	0.2	0.9037	0.1807	0.3685
4.0	0.2	0.3495	0.0699	0.1425

Posterior summaries are computed as before.

```
postmean <- round(sum(mu*post2),4)
postvar <- round(sum(mu^2*post2) - postmean^2,4)
pmulessthan3 <- table2 %>%
  filter(mu<3) %>%
  select(post2) %>%
  sum()
```

```r
knitr::kable(cbind(postmean, postvar, pmulessthan3),
             col.names = c("Posterior Mean",
                           "Posterior variance",
                           "P(mu < 3)"))
```

Posterior Mean	Posterior variance	P(mu < 3)
3.2496	0.219	0.1385

Consequently, Bayesian inference such as test of hypothesis and credible interval construction is based on the posterior distribution

Bayes inference for the normal mean using continuous prior

Suppose we have a random sample $Y_1, Y_2, \cdots, Y_n$ from $N(\mu, \sigma^2)$, where σ^2 is known. 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 for the normal mean

- Prior: $f(\mu) = 1$
- the posterior distribution is $N(\mu, \frac{\sigma^2}{n})$

Normal prior density for the normal mean

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

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

Hence, the posterior distribution of μ is

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

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

$$\mu^* = \frac{\sigma^2\mu_p + n\sigma_p^2\bar{y}}{\sigma^2 + n\sigma_p^2}$$

and

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

Alternatively, we can also express the posterior mean as

$$\mu^* = \left(\frac{\frac{1}{\sigma_p^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_p^2}}\right)\mu_p + \left(\frac{\frac{1}{\sigma_p^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_p^2}}\right)\bar{y}$$

Thus, it is cleear 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.

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

Example 3:

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. 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: $f(\mu) \approx exp\left[-\frac{(\mu-30)^2}{2(16)}\right]$

- Marie: $f(\mu) = 1$
- They take a random sample of 12 tilapia from the pond and calculated the sample mean to be $\bar{y} = 32cm$
- The likelihood: $y_i \sim N(\mu, 2^2) \implies \bar{y} \sim N(\mu, \frac{2^2}{12})$
- 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

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

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

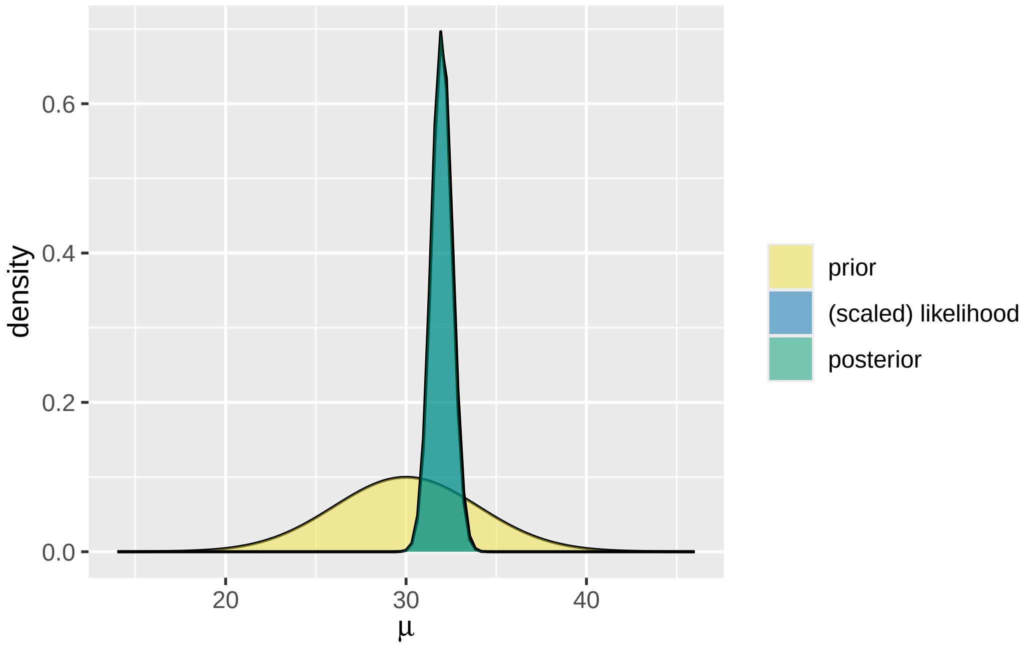
$$\mu^* = \frac{\sigma^2\mu_p + n\sigma_p^2\bar{y}}{\sigma^2 + n\sigma_p^2} = \frac{2^2(30) + 12(16)(32)}{2^2 + 12(16)} \approx 31.96$$

and

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

- 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)
```




```
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 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 demonsrated earlier, we can performed updating by the R function *normal_update()* nin 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 3 (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
- Simulation-based

```r
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%
30.87638 33.13072
```

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

```
    2.5%    97.5%
30.83487 33.07872
```

Bayesian one-sided hypothesis test about μ

- 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 $\alpha = 0.05$
- Simulation-based:

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

```
[1] 0.04305
```

- Let us test $H_0 : \mu \leq 31$ versus $H_1 : \mu > 31$ 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 $\alpha = 0.05$
- Simulation-based:

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

```
[1] 0.04637
```

Bayesian two-sided hypothesis test about μ

- Suppose we want to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$
- Recall that we have a continuous posterior, thus, the probability of any specific value of a continuous random variable always equals 0
- 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 $\alpha = 0.05$
- 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 tilapia
- 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
```

- Based on this Bayesian model, the expected length of a tilapia is 32 cm.