$X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_\theta$ $\qquad \hat{\theta}_n = MLE$

$$\ell(\theta; X) = \log f(X; \theta) = \log \prod_{i=1}^{n} f(x_i; \theta)$$
$$= \sum_{i=1}^{n} \log f(x_i; \theta)$$
$$= \sum_{i=1}^{n} \ell(\theta; x_i)$$

Score function: $\dfrac{d}{d\theta} \ell(\theta; X)$

Fisher information:

$$I_n(\theta) = \mathrm{Var}\left(\underline{\frac{d}{d\theta} \ell(\theta; X)}\right)$$

Standard error:

$$se(\hat{\theta}_n) \approx \frac{1}{\sqrt{I_n(\theta)}}$$

$$\hat{se}(\theta_n) = \frac{1}{\sqrt{I_n(\hat{\theta}_n)}}$$
$$= se(\hat{\theta}_n)$$

Asymptotic Normal:

$$\frac{\hat{\theta}_n - \theta}{\hat{se}(\theta_n)} \xrightarrow{D} \text{Normal}(0,1)$$

Pivotal statistic:

$$Z = \frac{\hat{\theta}_n - \theta}{\hat{se}(\theta_n)}$$ is approx. pivotal $N(0,1)$

large $\underline{n}$

Wald Test:

$H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$

$$z = \frac{\hat{\theta} - \theta_0}{\hat{se}(\theta)}$$

$$\text{p-value} = \Pr(|Z^*| \geq |z|)$$

$$Z^* \sim \text{Normal}(0,1)$$

MLE CI's

Two-sided CI $(1-\alpha)$ level is

$$\left(\hat{\theta} - |z_{\alpha/2}|\,\hat{se}(\theta),\ \hat{\theta} + |z_{\alpha/2}|\,\hat{se}(\theta)\right)$$

Delta Method

If $\hat{\theta}$ is MLE of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$.

Example: $X \sim \text{Binomial}(n, p)$

$\hat{p} = \frac{X}{n}$ is MLE for p

$n\hat{p}(1-\hat{p})$ is MLE for $np(1-p)$

Suppose $g(\cdot)$ is differentiable. Then

$$Var(g(\hat{\theta})) \approx g'(\theta)^2\, Var(\hat{\theta})$$

$$\Longrightarrow \hat{se}(g(\hat{\theta})) = |g'(\hat{\theta})|\,\hat{se}(\hat{\theta})$$

Example: $g(p) = p(1-p)$

$g'(p) = 1 - 2p$

$$\hat{se}(\hat{p}(1-\hat{p})) = |(1-2\hat{p})| \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Exponential Family Distribution (EFD)

A dist'n whose pmf or pdf has the following form:

$$f(x;\theta) = h(x) \exp\left\{ \sum_{k=1}^{d} \eta_k(\theta) T_k(x) - A(\eta) \right\}$$

θ is a vector of usual parameters

$T_k(x)$ sufficient statistics $\eta_k(\theta)$

η is a vector s.t. $\eta_k = \eta_k(\theta)$

$\eta_k(\theta)$ are called natural parameters

Example: Bernoulli

$$f(x;p) = p^x(1-p)^{1-x}$$

$$= \exp\{x\log(p) + (1-x)\log(1-p)\}$$

$$= \exp\{x\log\left(\frac{p}{1-p}\right) + \log(1-p)\}$$

$d = 1$

$\eta(p) = \log\left(\frac{p}{1-p}\right)$ $\qquad E[X] = p$

$T(x) = x$

$A(\eta) = \log(1-p)$

Example: Normal(μ, σ^2)

$\theta = (\mu, \sigma^2)$

$E[X] = \mu$

$E[X^2] = \sigma^2 + \mu^2$

$Var(X) = E[X^2] - E[X]^2 = \sigma^2$

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{ \frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 - \log(\sigma) - \frac{\mu^2}{2\sigma^2} \right\}$$

$$\eta(\mu,\sigma^2) = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right) \qquad d=2$$

$$\eta_1(\theta) = \frac{\mu}{\sigma^2}$$

$$\eta_2(\theta) = -\frac{1}{2\sigma^2}$$

$$T(x) = (x, x^2)$$

$$A(\eta) = \log(\sigma) + \frac{\mu^2}{2\sigma^2} = -\frac{1}{2}\log(-2\eta_2) - \frac{\eta_1^2}{4\eta_2}$$

Poisson is an EFD

Moments

k^{th} moment $E[X^k]$

$$E[T_k(X)] = \frac{\partial}{\partial \eta_k} A(\eta)$$

$$Var[T_k(X)] = \frac{\partial^2}{\partial \eta_k^2} A(\eta)$$

MLEs of $\eta_1, \eta_2, \dots, \eta_d$

MLE of η_k is the solution to

$$\frac{1}{n} \sum_{i=1}^{n} T_k(x_i) = \frac{\partial}{\partial \eta_k} A(\eta)$$

Dist'n	MLE	Std Err
Binomial (n, p)	$\hat{p} = X/n$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Normal (μ, σ^2)	$\hat{\mu} = \bar{X}$	$\hat{\sigma}/\sqrt{n}$
Poisson (λ)	$\hat{\lambda} = \bar{X}$	$\sqrt{\hat{\lambda}/n}$

For Normal(μ, σ^2) data, the MLE of σ^2 is:

$$\hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

t-distribution

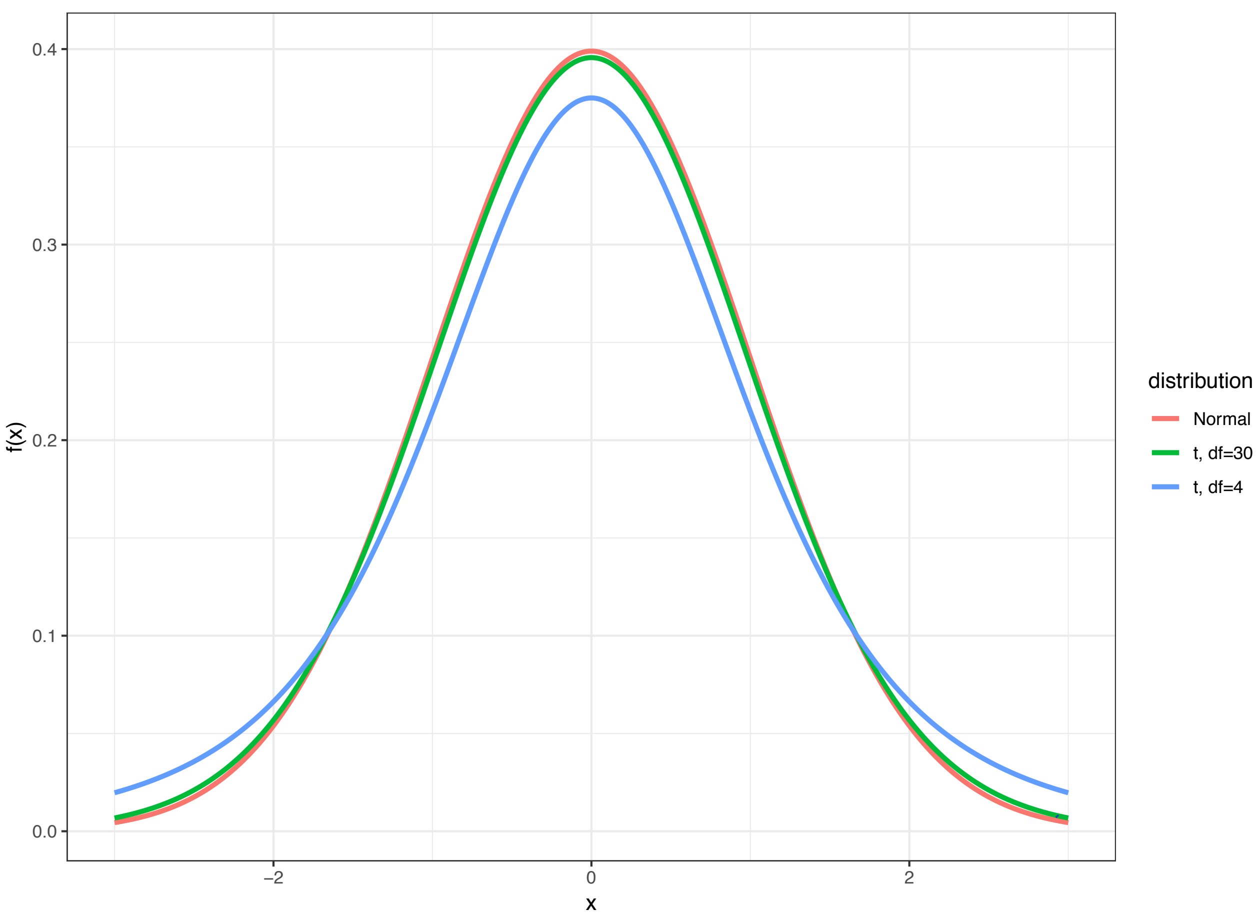
Allows us to return to the ideal scenario $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$ but now μ and σ^2 are unknown

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

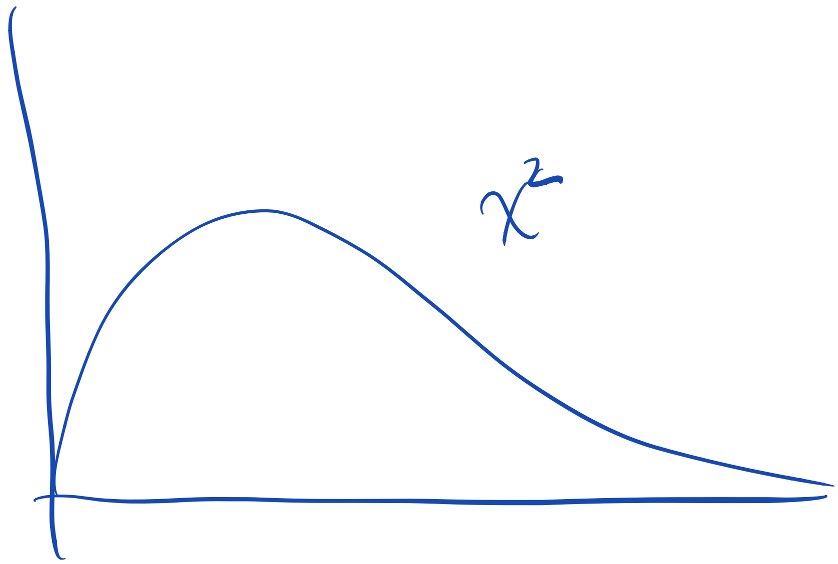
$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}$$

t vs Normal



$$\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\underbrace{\frac{(n-1)}{\sigma^2} S^2}_{W} \sim \chi^2_{n-1}$$



$\bar{X}$ and S^2 are independent

$(X_i - \bar{X})$

t_ν is defined by

$$\frac{Z}{\sqrt{W/\nu}}$$

$Z \sim \text{Normal}(0,1)$

$W \sim \chi^2_\nu$

Z and W are independent

t-percentiles

$$t_{0.025,4} \approx -2.78 \qquad \text{vs}$$

$$z_{0.025} = -1.96$$

Pivotal Statistic

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}$$

does not depend on μ or σ^2

Just like last week

$\sigma^2 \leftarrow S^2$

Normal$(0,1) \leftarrow t_{n-1}$

Comparing two populations

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_{\theta_1}$

$Y_1, Y_2, \ldots, Y_m \overset{iid}{\sim} F_{\theta_2}$

Recall that:

$E[X-Y] = E[X] - E[Y]$

If independent, then:

$Var(X-Y) = Var(X) + Var(Y)$

Let's say $\hat{\theta}_1$ is MLE from X

and $\hat{\theta}_2$ is MLE from Y

$\hat{\theta}_1 \sim Normal(\theta_1, \hat{se}^2(\theta_1))$

$\hat{\theta}_2 \sim Normal(\theta_2, \hat{se}^2(\theta_2))$

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\hat{se}^2(\hat{\theta}_1) + \hat{se}^2(\hat{\theta}_2)}} \sim Normal(0, 1)$$

$$Z = \frac{X - Y - (E[X] - E[Y])}{\sqrt{Var(X) + Var(Y)}}$$

$$E[Z] = 0$$

$$Var(Z) = 1$$

Wald test: $\theta_1 = \theta_2$

$$H_0: \overbrace{\theta_1 - \theta_2 = 0} \quad vs$$

$$H_1: \theta_1 - \theta_2 \neq 0$$

Apprx $(1-\alpha)$ CI on $\theta_1 - \theta_2$

$$\left(\hat{\theta}_1 - \hat{\theta}_2 - |z_{\alpha/2}|\,\hat{se},\ \hat{\theta}_1 - \hat{\theta}_2 + |z_{\alpha/2}|\,\hat{se}\right)$$

$$\hat{se} = \sqrt{\hat{se}^2(\hat{\theta}_1) + \hat{se}^2(\hat{\theta}_2)}$$

Inference in R

BSDA Package

```
> install.packages("BSDA")
```

```
> library(BSDA)
> str(z.test)
function (x, y = NULL, alternative = "two.sided", mu = 0, sigma.x = NULL,
    sigma.y = NULL, conf.level = 0.95)
```

Example: Poisson

Apply `z.test()`:

```
> set.seed(210)
```

```
> n <- 40
> lam <- 14
> x <- rpois(n=n, lambda=lam)
> lam.hat <- mean(x)
> stddev <- sqrt(lam.hat)
> z.test(x=x, sigma.x=stddev, mu=lam)

    One-sample z-Test

data:  x
z = 0.41885, p-value = 0.6753
alternative hypothesis: true mean is not equal to 14
95 percent confidence interval:
 13.08016 15.41984
sample estimates:
mean of x
    14.25
```

Direct Calculations

Confidence interval:

```
> lam.hat <- mean(x)
> lam.hat
[1] 14.25
> stderr <- sqrt(lam.hat)/sqrt(n)
> lam.hat - abs(qnorm(0.025)) * stderr # lower bound
[1] 13.08016
> lam.hat + abs(qnorm(0.025)) * stderr # upper bound
[1] 15.41984
```

Hypothesis test:

```
> z <- (lam.hat - lam)/stderr
> z # test statistic
[1] 0.4188539
> 2 * pnorm(-abs(z)) # two-sided p-value
[1] 0.6753229
```

Commonly Used Functions

R has the following functions for doing inference on some of the distributions we have considered.

- Normal: `t.test()`
- Binomial: `binomial.test()` or `prop.test()`
- Poisson: `poisson.test()`

These perform one-sample and two-sample hypothesis testing and confidence interval construction for both the one-sided and two-sided cases.

About These Functions

- We covered a convenient, unified MLE framework that allows us to better understand how confidence intervals and hypothesis testing are performed
- However, this framework requires large sample sizes and is not necessarily the best method to apply in all circumstances
- The above R functions are versatile functions for analyzing Normal, Binomial, and Poisson distributed data (or approximations thereof) that use much broader theory and methods than we have covered
- However, the arguments these functions take and the ouput of the functions are in line with the framework that we have covered

Normal Data: "Davis" Data Set

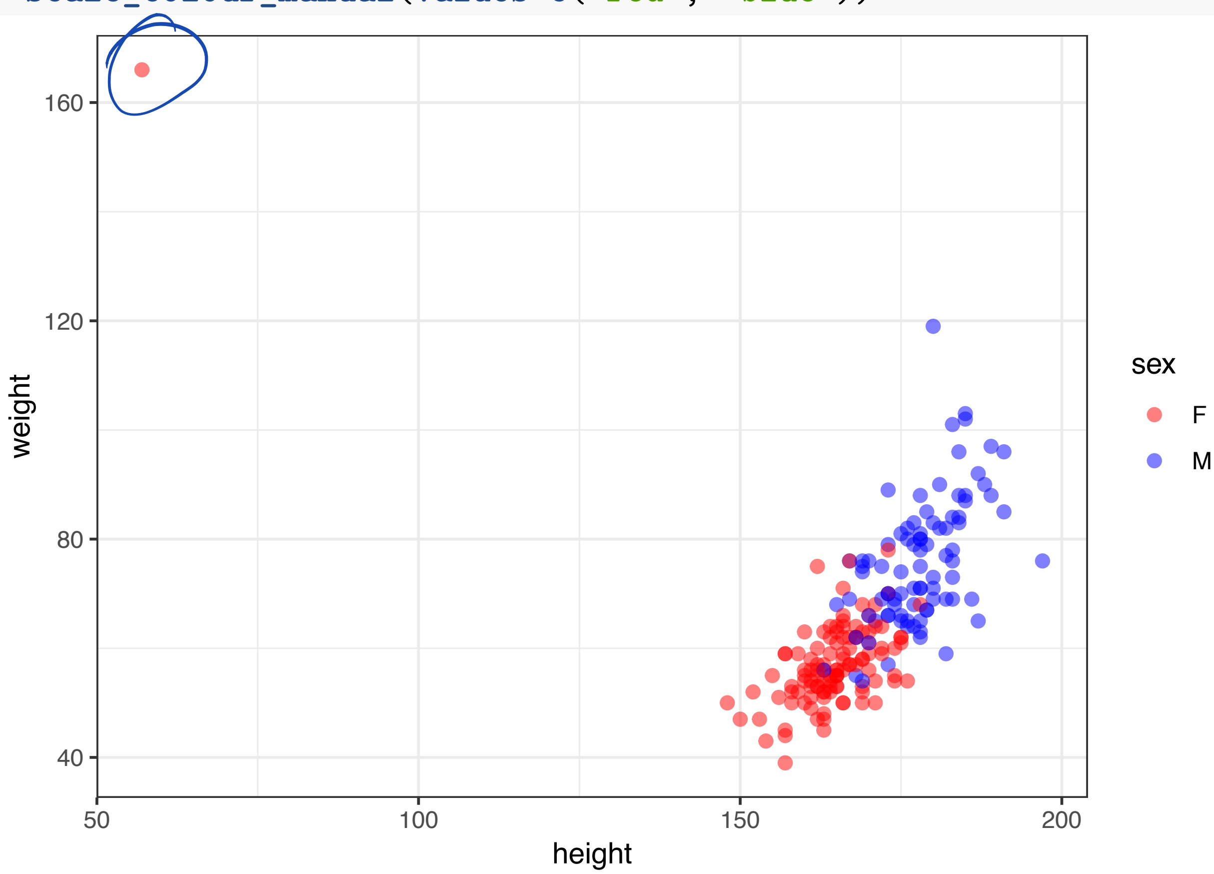
```
> library("car")
> data("Davis")
```

```
> htwt <- tbl_df(Davis)
> htwt
# A tibble: 200 x 5
    sex   weight height repwt repht
    <fct>  <int>  <int> <int> <int>
  1 M         77    182    77   180
  2 F         58    161    51   159
  3 F         53    161    54   158
```

```
 4 M           68     177     70    175
 5 F           59     157     59    155
 6 M           76     170     76    165
 7 M           76     167     77    165
 8 M           69     186     73    180
 9 M           71     178     71    175
10 M           65     171     64    170
# ... with 190 more rows
```

Height vs Weight

```
> ggplot(htwt) +
+ geom_point(aes(x=height, y=weight, color=sex), size=2, alpha=0.5) +
+ scale_colour_manual(values=c("red", "blue"))
```



An Error?

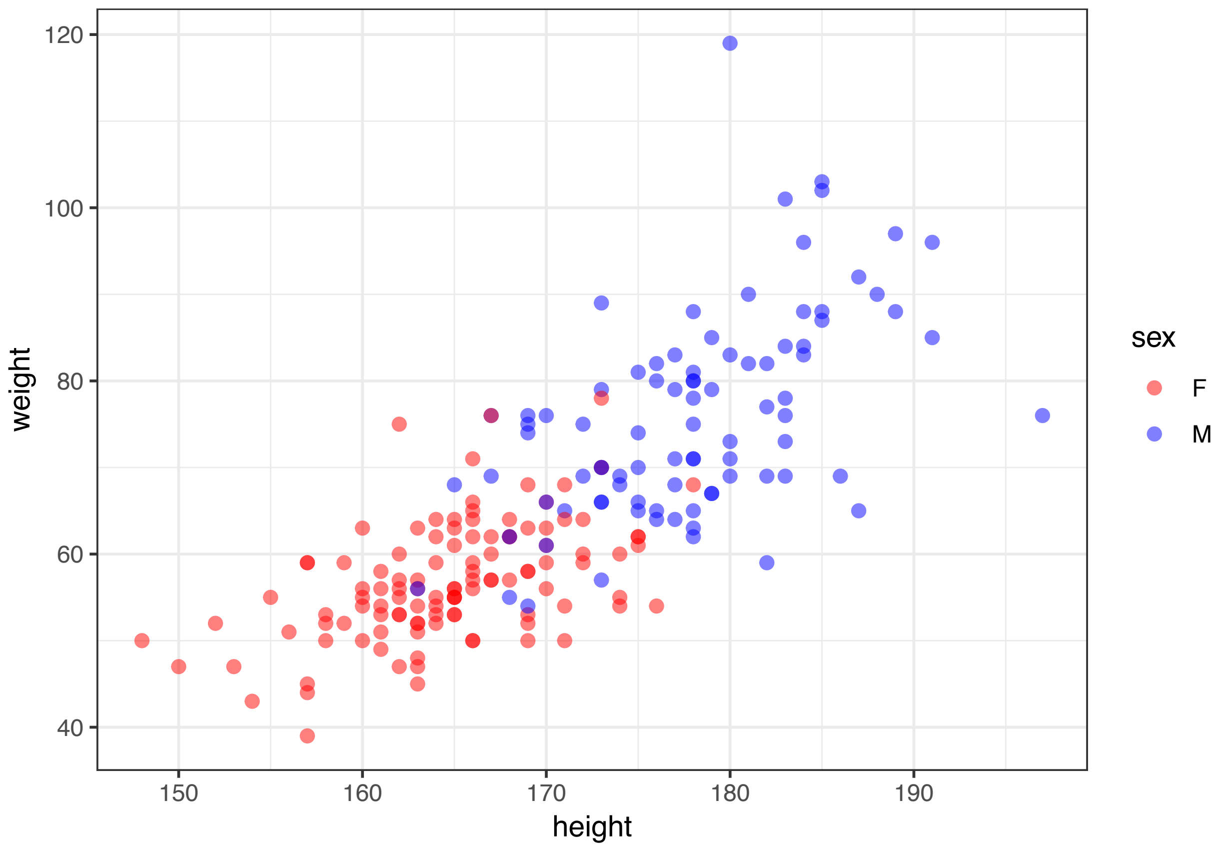
```
> which(htwt$height < 100)
[1] 12
> htwt[12,]
# A tibble: 1 x 5
```

```
  sex   weight height repwt repht
  <fct>  <int>  <int> <int> <int>
1 F        166     57    56   163
```

```
> htwt[12,c(2,3)] <- htwt[12,c(3,2)]
```

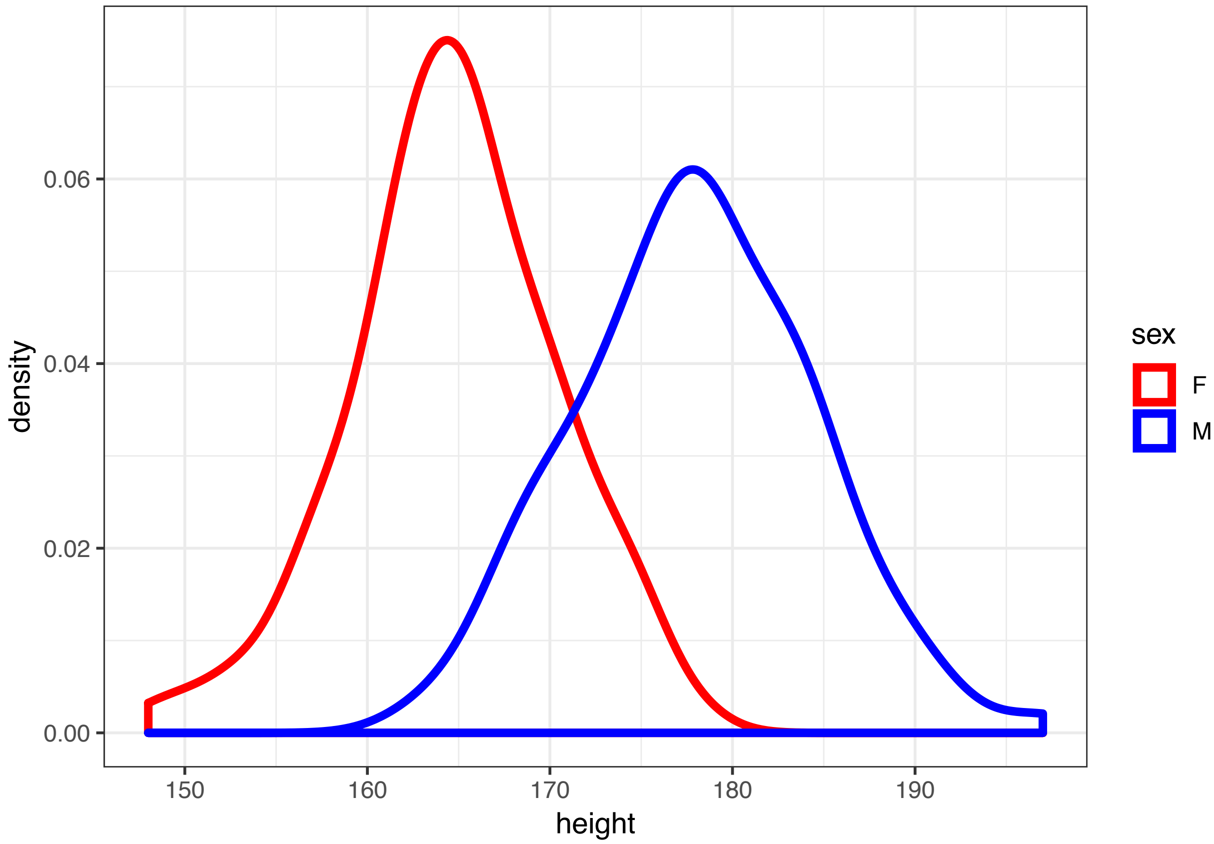
Updated Height vs Weight

```
> ggplot(htwt) +
+   geom_point(aes(x=height, y=weight, color=sex), size=2, alpha=0.5) +
+   scale_color_manual(values=c("red", "blue"))
```



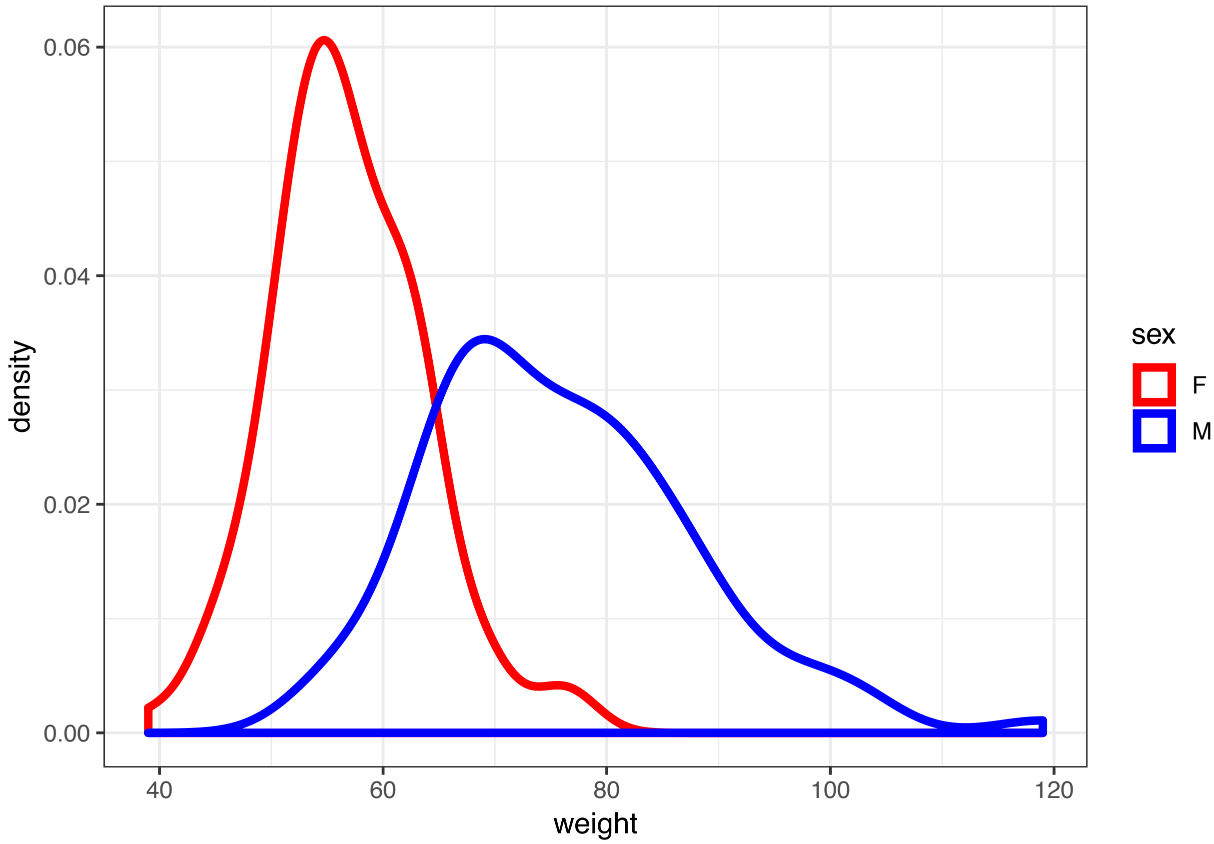
Density Plots of Height

```
> ggplot(htwt) +
+   geom_density(aes(x=height, color=sex), size=1.5) +
+   scale_color_manual(values=c("red", "blue"))
```



Density Plots of Weight

```
> ggplot(htwt) +
+   geom_density(aes(x=weight, color=sex), size=1.5) +
+   scale_color_manual(values=c("red", "blue"))
```



`t.test()` Function

From the help file…

Usage

```
t.test(x, ...)

## Default S3 method:
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)

## S3 method for class 'formula'
t.test(formula, data, subset, na.action, ...)
```

Two-Sided Test of Male Height

```
> m_ht <- htwt %>% filter(sex=="M") %>% select(height)
> testresult <- t.test(x = m_ht$height, mu=177)
```

```
> class(testresult)
[1] "htest"
> is.list(testresult)
[1] TRUE
```

Output of t.test()

```
> names(testresult)
[1] "statistic"   "parameter"   "p.value"     "conf.int"    "estimate"
[6] "null.value"  "alternative" "method"      "data.name"
> testresult

	One Sample t-test

data:  m_ht$height
t = 1.473, df = 87, p-value = 0.1443
alternative hypothesis: true mean is not equal to 177
95 percent confidence interval:
 176.6467 179.3760
sample estimates:
mean of x
 178.0114
```

Tidying the Output

```
> library(broom)
> tidy(testresult)
# A tibble: 1 x 8
  estimate statistic p.value parameter conf.low conf.high method
     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
1     178.      1.47   0.144        87     177.      179. One S~
# ... with 1 more variable: alternative <chr>
```

Two-Sided Test of Female Height

```
> f_ht <- htwt %>% filter(sex=="F") %>% select(height)
> t.test(x = f_ht$height, mu = 164)

	One Sample t-test
```

```
data:  f_ht$height
t = 1.3358, df = 111, p-value = 0.1844
alternative hypothesis: true mean is not equal to 164
95 percent confidence interval:
 163.6547 165.7739
sample estimates:
mean of x
 164.7143
```

Difference of Two Means

```
> t.test(x = m_ht$height, y = f_ht$height)

	Welch Two Sample t-test

data:  m_ht$height and f_ht$height
t = 15.28, df = 174.29, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 11.57949 15.01467
sample estimates:
mean of x mean of y
 178.0114  164.7143
```

Test with Equal Variances

```
> htwt %>% group_by(sex) %>% summarize(sd(height))
# A tibble: 2 x 2
  sex   `sd(height)`
  <fct>        <dbl>
1 F             5.66
2 M             6.44
> t.test(x = m_ht$height, y = f_ht$height, var.equal = TRUE)

	Two Sample t-test

data:  m_ht$height and f_ht$height
t = 15.519, df = 198, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 11.60735 14.98680
sample estimates:
mean of x mean of y
 178.0114  164.7143
```

Paired Sample Test (v. 1)

First take the difference between the paired observations. Then apply the one-sample t-test.

```
> htwt <- htwt %>% mutate(diffwt = (weight - repwt),
+                          diffht = (height - repht))
> t.test(x = htwt$diffwt) %>% tidy()
# A tibble: 1 x 8
  estimate statistic p.value parameter conf.low conf.high method
     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
1  0.00546    0.0319   0.975       182   -0.332     0.343 One S~
# ... with 1 more variable: alternative <chr>
> t.test(x = htwt$diffht) %>% tidy()
# A tibble: 1 x 8
  estimate statistic  p.value parameter conf.low conf.high method
     <dbl>     <dbl>    <dbl>     <dbl>    <dbl>     <dbl> <chr>
1     2.08      13.5 2.64e-29       182     1.77      2.38 One S~
# ... with 1 more variable: alternative <chr>
```

Paired Sample Test (v. 2)

Enter each sample into the `t.test()` function, but use the `paired=TRUE` argument. This is operationally equivalent to the previous version.

```
> t.test(x=htwt$weight, y=htwt$repwt, paired=TRUE) %>% tidy()
# A tibble: 1 x 8
  estimate statistic p.value parameter conf.low conf.high method
     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
1  0.00546    0.0319   0.975       182   -0.332     0.343 Paire~
# ... with 1 more variable: alternative <chr>
> t.test(x=htwt$height, y=htwt$repht, paired=TRUE) %>% tidy()
# A tibble: 1 x 8
  estimate statistic  p.value parameter conf.low conf.high method
     <dbl>     <dbl>    <dbl>     <dbl>    <dbl>     <dbl> <chr>
1     2.08      13.5 2.64e-29       182     1.77      2.38 Paire~
# ... with 1 more variable: alternative <chr>
> htwt %>% select(height, repht) %>% na.omit() %>%
+   summarize(mean(height), mean(repht))
# A tibble: 1 x 2
  `mean(height)` `mean(repht)`
           <dbl>         <dbl>
1           171.          168.
```

The Coin Flip Example

I flip it 20 times and it lands on heads 16 times.

1. My data is $x = 16$ heads out of $n = 20$ flips.
2. My data generation model is $X \sim \text{Binomial}(20, p)$.
3. I form the statistic $\hat{p} = 16/20$ as an estimate of p.

Let's do hypothesis testing and confidence interval construction on these data.

`binom.test()`

```
> str(binom.test)
function (x, n, p = 0.5, alternative = c("two.sided", "less", "greater"),
    conf.level = 0.95)
> binom.test(x=16, n=20, p = 0.5)

	Exact binomial test

data:  16 and 20
number of successes = 16, number of trials = 20, p-value = 0.01182
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.563386 0.942666
sample estimates:
probability of success
                   0.8
```

`alternative = "greater"`

Tests $H_0 : p \leq 0.5$ vs. $H_1 : p > 0.5$.

```
> binom.test(x=16, n=20, p = 0.5, alternative="greater")

	Exact binomial test

data:  16 and 20
number of successes = 16, number of trials = 20, p-value =
0.005909
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.5989719 1.0000000
sample estimates:
probability of success
                   0.8
```

alternative = "less"

Tests $H_0 : p \geq 0.5$ vs. $H_1 : p < 0.5$.

```
> binom.test(x=16, n=20, p = 0.5, alternative="less")

	Exact binomial test

data:  16 and 20
number of successes = 16, number of trials = 20, p-value = 0.9987
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.9286461
sample estimates:
probability of success
                   0.8
```

prop.test()

This is a "large n" inference method that is very similar to our z-statistic approach.

```
> str(prop.test)
function (x, n, p = NULL, alternative = c("two.sided", "less", "greater"),
    conf.level = 0.95, correct = TRUE)
> prop.test(x=16, n=20, p=0.5)

	1-sample proportions test with continuity correction

data:  16 out of 20, null probability 0.5
X-squared = 6.05, df = 1, p-value = 0.01391
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5573138 0.9338938
sample estimates:
  p
0.8
```

An Observation

```
> p <- binom.test(x=16, n=20, p = 0.5)$p.value
> binom.test(x=16, n=20, p = 0.5, conf.level=(1-p))

	Exact binomial test

data:  16 and 20
number of successes = 16, number of trials = 20, p-value = 0.01182
```

```
alternative hypothesis: true probability of success is not equal to 0.5
98.81821 percent confidence interval:
 0.5000000 0.9625097
sample estimates:
probability of success
                   0.8
```

Exercise: Figure out what happened here.

Example: RNA-Seq

RNA-Seq gene expression was measured for p53 lung tissue in 12 healthy individuals and 14 individuals with lung cancer.

The counts were given as follows.

Healthy: 82 64 66 88 65 81 85 87 60 79 80 72 — $Pois(\lambda_1)$

Cancer: 59 50 60 60 78 69 70 67 72 66 66 68 54 62 — $Pois(\lambda_2)$

λ_1 / λ_2

It is hypothesized that p53 expression is higher in healthy individuals. Test this hypothesis, and form a 99% CI.

$H_1 : \lambda_1 \neq \lambda_2$

```
> healthy <- c(82, 64, 66, 88, 65, 81, 85, 87, 60, 79, 80, 72)
> cancer <- c(59, 50, 60, 60, 78, 69, 70, 67, 72, 66, 66, 68,
+              54, 62)
```

Poisson Data: `poisson.test()`

```
> str(poisson.test)
function (x, T = 1, r = 1, alternative = c("two.sided", "less", "greater"),
    conf.level = 0.95)
```

From the help:

```
Arguments

x    number of events. A vector of length one or two.

T    time base for event count. A vector of length one or two.

r    hypothesized rate or rate ratio

alternative  indicates the alternative hypothesis and must be one of
"two.sided", "greater" or "less". You can specify just the initial letter.

conf.level  confidence level for the returned confidence interval.
```

```
> poisson.test(x=c(sum(healthy), sum(cancer)), T=c(12,14),
+              conf.level=0.99)

    Comparison of Poisson rates

data:  c(sum(healthy), sum(cancer)) time base: c(12, 14)
count1 = 909, expected count1 = 835.38, p-value = 0.0005739
```

```
alternative hypothesis: true rate ratio is not equal to 1
99 percent confidence interval:
 1.041626 1.330051
sample estimates:
rate ratio
  1.177026
```

$H_1 : \lambda_1 < \lambda_2$

```
> poisson.test(x=c(sum(healthy), sum(cancer)), T=c(12,14),
+               alternative="less", conf.level=0.99)

    Comparison of Poisson rates

data:  c(sum(healthy), sum(cancer)) time base: c(12, 14)
count1 = 909, expected count1 = 835.38, p-value = 0.9998
alternative hypothesis: true rate ratio is less than 1
99 percent confidence interval:
 0.000000 1.314529
sample estimates:
rate ratio
  1.177026
```

$H_1 : \lambda_1 > \lambda_2$

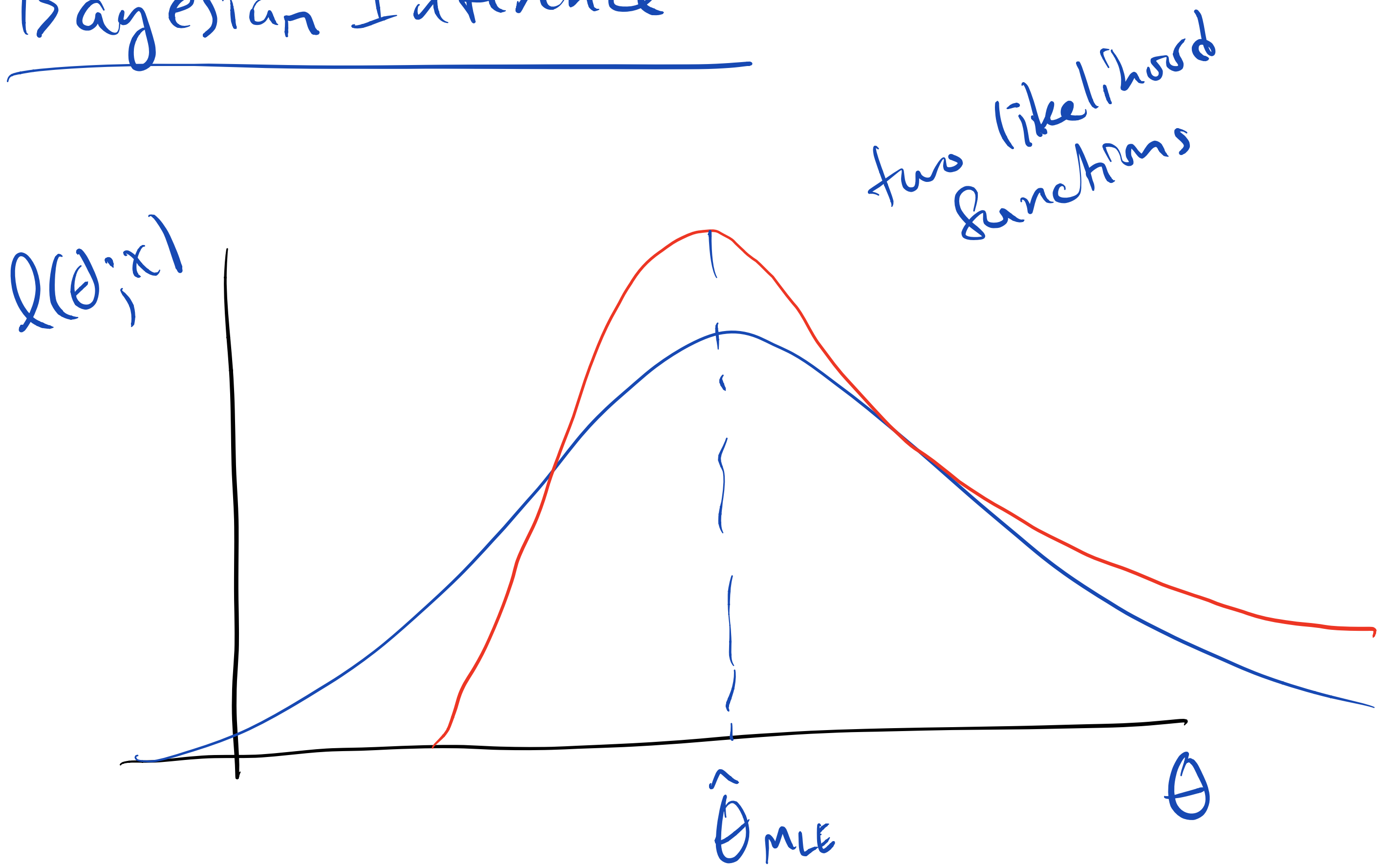
$H_0: \lambda_1 \leq \lambda_2$ vs $H_1: \lambda_1 > \lambda_2$

```
> poisson.test(x=c(sum(healthy), sum(cancer)), T=c(12,14),
+               alternative="greater", conf.level=0.99)

    Comparison of Poisson rates

data:  c(sum(healthy), sum(cancer)) time base: c(12, 14)
count1 = 909, expected count1 = 835.38, p-value = 0.0002881
alternative hypothesis: true rate ratio is greater than 1
99 percent confidence interval:
 1.053921      Inf
sample estimates:
rate ratio
  1.177026
```

Question

Which analysis is the more informative and scientifically correct one, and why?

Bayesian Inference



Frequentist probability vs.
subjective Bayesian probability

In Bayesian inference, the parameter(s) have a "prior" probability distribution.

Update this prior by calculating the "posterior" dist'n of the parameter(s) conditional on the data using Bayes theorem.

$$\left(X_1, X_2, \dots, X_n \mid \theta \overset{iid}{\sim} F_\theta \right.$$

$$\theta \sim F_\tau \longleftarrow \text{prior distribution}$$

$$\theta \mid X_1, X_2, \dots, X_n \sim \ ? \quad \text{posterior dist'n}$$

Example: Prior is $P \sim \text{Uniform}(0,1)$

$$X \mid P = p \sim \text{Binomial}(n, p)$$

$$f(p) = 1 \qquad 0 \le p \le 1$$

$$f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(p \mid x) = \frac{f(x \mid p) f(p)}{f(x)}$$

$$f(x) = \int f(x \mid p) f(p) \, dp$$

$$f(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta) f(\theta)}{f(\mathbf{x})}$$

$$f(x) = \int f(x|\theta^*) f(\theta^*) d\theta^*$$

$$f(\theta|X) \propto f(X|\theta) f(\theta)$$

$$= L(\theta|X) f(\theta)$$

$$E[\theta|X] = \int \theta f(\theta|x) dx$$

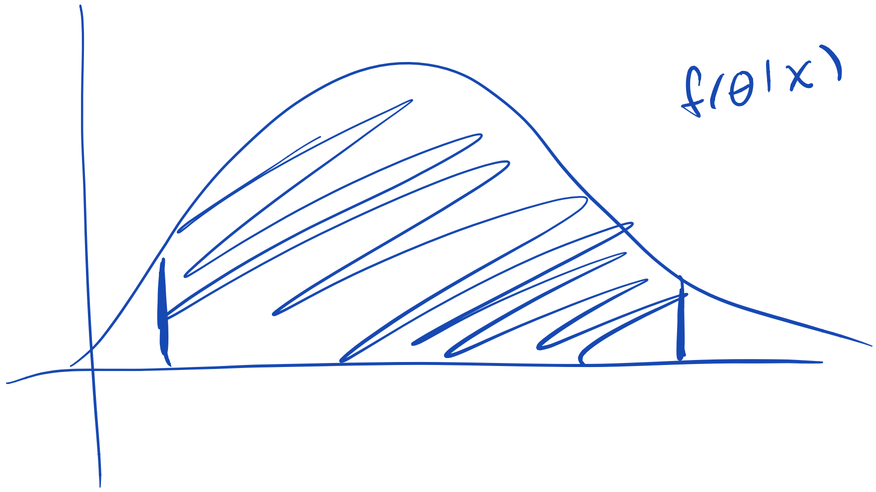
posterior expectation

$$\propto \int \theta \underline{L(\theta|x)}\, \underline{f(\theta)}\, dx$$

Bayesian Point Estimates of θ

1. $E[\theta|X]$ posterior expectation
2. $\underset{\theta}{\arg\max}\, f(\theta|x)$ MAP maximum a posteriori probability
3. Percentiles from $f(\theta|x)$

Posterior Interval



$1-\alpha$ posterior interval, identify C_ℓ and C_u so that

$$\int_{C_\ell}^{C_u} f(\theta|x)\,d\theta = 1-\alpha$$

Loss Functions

$\mathcal{L}(\theta, \tilde{\theta})$ quantifies the loss in estimating θ by $\hat{\theta}$

$$\mathcal{L}(\theta, \tilde{\theta}) = (\theta - \tilde{\theta})^2$$

$$\mathcal{L}(\theta, \tilde{\theta}) = |\theta - \tilde{\theta}|$$

Bayes Risk

$$E[\mathcal{L}(\theta, \tilde{\theta}) \mid \underline{X}]$$

$\underline{X}$ data

$$= \int \mathcal{L}(\theta, \tilde{\theta}) f(\theta \mid x)\, d\theta$$

Bayes Estimator

The estimate that minimizes Bayes risk

$E[\theta \mid \underline{X}]$ minimizes $\mathcal{L}(\theta, \tilde{\theta}) = (\theta - \tilde{\theta})^2$

median of $f(\theta \mid x)$ minimizes
$\mathcal{L}(\theta, \tilde{\theta}) = |\theta - \tilde{\theta}|$

How to determine prior?

Conjugate prior: this is a prior for d.g.d. so that the posterior is of the same distribution as the prior

Example. $X \mid P=p \sim \text{Binomial}(n, p)$

$P \sim \text{Beta}(\alpha, \beta)$

$$f(p \mid x) \propto L(p \mid x) f(p)$$

$$= p^{x}(1-p)^{n-x} p^{\alpha-1}(1-p)^{\beta-1}$$

$$= p^{\alpha-1+x}(1-p)^{\beta-1+(n-x)}$$

$$\propto \text{Beta}(\alpha+x, \beta+(n-x))$$

$$E[p \mid x] = \frac{\alpha+x}{\alpha+\beta+n}$$

$$E(p) = \frac{\alpha}{\alpha+\beta}$$

$$\hat{p}_{MLE} = \frac{X}{n}$$

Example Normal - Normal for μ

Jeffreys Prior

If $f(\theta)$ the prior pdf is such that $f(\theta) \propto \sqrt{I(\theta)}$ then the posterior will be invariant to transformations of θ

Examples:

Normal(μ, σ^2), σ^2 known: $f(\mu) \propto 1$

Normal (μ, σ^2), μ known: $f(\sigma) \propto \frac{1}{\sigma}$

Poisson(λ): $f(\lambda) \propto \frac{1}{\sqrt{\lambda}}$

Bernoulli(p): $f(p) \propto \frac{1}{\sqrt{p(1-p)}}$

$X_1, X_2, \ldots, X_n \mid \mu \sim \text{Normal}(\mu, \sigma^2)$

σ^2 is known

Prior $f(\mu) \propto 1$ $\quad \int f(\mu)\, d\mu = \infty$

$$f(\mu \mid \underline{x}) \propto L(\mu \mid \underline{x})\, f(\mu)$$

$$\sim \text{Normal}(\bar{x}, \sigma^2/n)$$