Review of Basic Statistical Procedures

Basic Statistical Inference

Let $X_1, \cdots, X_n$ be a random sample with pdf $f(\theta)$, where either f or θ or both may be unknown.

Inference:

- Estimation
- Hypothesis Testing

Approaches:

- Frequentist
- Bayesian

Parameter Estimation

- The Maximum Likelihood Method

  Let $X_, \cdots, X_n$ be a random sample from $f(x, \theta)$. The *likelihood function* of θ is defined to be

$$L(\theta) = \Pi_i^n f(x_i; \theta)$$

  Under regularity conditions, $\hat{\theta}$, the MLE, is approximately normal, with mean θ and variance $\frac{1}{nI(\theta)}$, where

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} \log f(x, \theta)\right]^2$$

- The Method of Moments

Define the k'th *moment* of $f(x,\theta)$ as
$\mu_k = E[X^k]$.
The corresponding *sample moments* are given as

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Let $\theta_j = h_j(\mu_1, \cdots, \mu_s)$ be parameters of interest.
The method of moments estimates of θ_j

$$\hat{\theta}_j = h_j(\hat{\mu}_1, \cdots \hat{\mu}_s)$$

Unbiasedness

An estimator $\hat{\theta}$ is said to be unbiased if $E[\hat{\theta}] = \theta$.

An estimator is said to be uniformly minimum variance unbiased (UMVU) if it has the smallest variance in the class of unbiased estimators uniformly in θ.

Consistency

An estimator $\hat{\theta}_n$ of θ is said to be *consistent* if it approaches θ for large n. More formally, for any $\epsilon > 0$,

$$Pr[|\ \hat{\theta}_n - \theta\ | > \epsilon] \to 0,\ n \to \infty$$

Method of moments estimators and most mle's are consistent.

Efficiency

Let $\tilde{\theta}$ and $\hat{\theta}$ be possible estimates of θ. The *efficiency* of $\hat{\theta}$ relative to $\tilde{\theta}$ is defined to be

$$\epsilon(\hat{\theta}, \tilde{\theta}) = \frac{Var(\tilde{\theta})}{Var(\hat{\theta})}$$

Cramer-Rao Inequality. Let $X_1, \cdots, X_n)$ be i.i.d. with p.d.f. $f(x;\theta)$. Let $T = T(X_, \cdots, X_n)$ be an unbiased estimator of θ. Then under regularity conditions

$$Var(T) \geq \frac{1}{nI(\theta)}$$

An unbiased estimator whose variance attains the lower bound is said to be *efficient*.

Sufficiency

A statistic $T = T(X_1, \cdots, X_n)$ is said to be sufficient for θ if the conditional distribution of the sample given $T = t$ is independent of θ, i.e., T contains all the information about θ.

A necessary and sufficient condition for T to be sufficient for θ is given by the *factorization theorem*

$$f(x_1, \cdots, x_n; \theta) = g[T, \theta] h(x_1, \cdots, x_n)$$

If there is a sufficient statistic for θ then the mle is also a function of T

Rao-Blackwell Theorem. Let $\hat{\theta}$ be an etsimator of θ, with $E[\hat{\theta}^2] < \infty$, and put $\tilde{\theta} = E[\hat{\theta} \mid T]$. Then

$$E(\tilde{\theta} - \theta)^2 \leq E(\hat{\theta} - \theta)^2$$

with strict equality when $\tilde{\theta} = \hat{\theta}$.

In *complete* families of distributions, if there is an unbiased estimator that is a function of the sufficient statistic, then one should look no further, i.e., that estimator is the unique unbiased minimum variance estimator.

Lehmann-Scheffe theorm.

Hypothesis Testing

Suppose we are interested in determining whether $\theta \epsilon \Theta_1$ versus $\theta \epsilon \Theta_2$, where Θ_1 and Θ_2 are disjoint subsets of the parameter space.

Suppose $X_1, \cdots, X_n$ is a random sample from $f(x, \theta)$, the hypotheses may be written as

$$H_0 : \theta \epsilon \Theta_0$$

$$H_1 : \theta \epsilon \Theta_1.$$

- Type I error occurs when H_o is falsely rejected

  Type II: When H_o is not rejected when it should be.

The probability of Type I error is known as *level of significance.*

Suppose H_0 specifies that the pdf of the data is $f_\circ(x)$, and H_1 claims $f_1(x)$.

Neyman-Pearson Lemma. Let ϕ^* be a test that accepts when

$$\frac{f_\circ(x)}{f_1(x)} > c$$

with significance level α^*. Let ϕ be another test of level $\alpha \leq \alpha^*$. Then the power of ϕ^* is greater than or equal to that of ϕ.

For testing composite hypotheses, $H_o : \theta \epsilon \Theta_o$ vs $H_1 : \theta \epsilon \Theta_1$, the *generalized likelihood ratio tests* may be constructed:

$$\Lambda = \frac{\sup_{\theta \epsilon \Theta_o}[L(\theta)]}{\sup_{\theta \epsilon \Theta_1}[L(\theta)]}$$

where Θ_0 and Θ_1 are partitions of the parameter space Θ.

In application, one uses the following formulation

$$\Lambda = \frac{\sup_{\theta \epsilon \Theta_o}[L(\theta)]}{\sup_{\theta \epsilon \Theta}[L(\theta)]}$$

The likelihood ratio test rejects for small values of Λ, or large values of $-log\Lambda$.

Given a test statistic $T = T(X_1, \cdots, X_n; \theta_o)$, and the observed value t_{obs}, the *p-value* is defined as the probability of observing a value of T as extreme as or more extreme than t_{obs}. Small p-values are evidence against H_o.

Confidence Intervals

Let $L = L(X_1, \cdots, X_n)$ and $U = U(X_1, \cdots, X_n)$ be statistics, such that $Pr[L \leq \theta \leq U] = 1-\alpha$. Then the interval $[L, U]$ is called a $100(1-\alpha)\%$ confidence interval for θ.

Assumptions of Classical Statistical Inference

- Independence
- Normality (or other specified distribution)

Validation of Assumptions:

- Histograms, qqnorm, etc., for normality
- Box-plots, stem-and-leaf diagrams for outlier detection
- Simple time series plot for serial correlations: including trends and cycles.

When distribution of data skewed, confidence intervals tend to be large, on the average, and p-values may be inflated.

Nonparametric methods do not make explicit assumptions about underlying distributions.

Most commonly used nonparametric methods are based on rank transformations of the observations.

Example: Cedar-apple rust data:

- Disease that affects apple trees, with symptom, rust-colored spots on apple leaves. Study conducted to study cause.
- In the first year of experiment the number of affected leaves on a random sample of 8 trees was counted on Variety 1.
- Normally, the average number of affected leaves is at most 25.

Data:

Tree	Count
1	38
2	10
3	84
4	36
5	50
6	35
7	73
8	48

Does the data suggest the apple trees are affected?

Standard Tests for Location

One-Sample Problem

Let $X_1, \cdots, X_n$ be a random sample from a distribution with meand μ and variance σ^2. We wish to test the hypothesis

$$H_0 : \mu = \mu_o$$

vs.

$$H_1 : \mu > \mu_o$$

When the distribution is normal, and σ^2 unknown (the usual case), a procedure that is commonly used is the Student's t-test, given by

$$T = \frac{\sqrt{n}(\bar{X} - \mu_o)}{S}$$

which under the null has a t_{n-1} distribution.

Example: Cedar-apple rust data:

```
> variety1
[1] 38 10 84 36 50 35 73 48

> t.test(variety1,mu=25)

        One Sample t-test

data:  variety1
t = 2.651, df = 7, p-value = 0.03289

alternative hypothesis: true mean is not equal to 25

95 percent confidence interval:
 27.34963 66.15037
sample estimates:
mean of x:    46.75
```

If EDA shows that the data is non-normal, then the *Wilcoxon signed-rank* test may be used.

- Let $D_i = X_i - \mu_o$
- Rank $\mid D_1 \mid, \cdots, \mid D_n \mid$

- Let T_+ and T_- be the sums of the ranks assigned to the positive and negative $d_i's$, respectively. Clearly, if H_0 is true, $T_+ \approx T_-$.

Further, under H_o,

$$E[T_+] = \frac{n(n+1)}{4}$$

and

$$Var(T_+) = \frac{n(n+1)(2n+1)}{24}$$

- The Wilcoxon signed-rank test rejects for large values of T_+, or when

$$Z = \frac{T_+ - E[T_+ \mid H_o]}{\sqrt{Var(T_+ \mid H_o)}}$$

exceeds a standard normal critical point.

- The test requires the underlying distribution to be symmetric.

Example: Cedar-apple rust data (cont'd):

- Disease that affects apple trees, with symptom, rust-colored spots on apple leaves. Study conducted to study cause.
- In the first year of experiment the number of affected leaves on a random sample of 8 trees from Variety 1 and a random sample of 7 trees from Variety 2 counted.

Data:

Variety 1		Variety 2	
Tree	Count	Tree	Count
1	38	1	27
2	10	2	28
3	84	3	57
4	36	4	66
5	50	5	77
6	35	6	49
7	73	7	62
8	48		

Does the data suggest there is a difference between the two varieties in the mean number of affected leaves?

Two-Sample Problem

Let $X_1, \cdots X_n$, and $Y_1, \cdots Y_m$, be independent random samples from distributions with respective means, μ_1 and μ_2, and variances, σ_1^2 and σ_2^2.

We are interested in the hypothesis

$$H_o : \mu_1 - \mu_2 = \triangle_o$$

vs

$$H_1 : \mu_1 - \mu_2 > \triangle_o$$

When both distributions are normal, with unkown, but equal, variance σ^2, the two-sample t-test is given by

$$T = \frac{(\bar{X} - \bar{Y}) - \triangle_0)}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}}$$

where the pooled variance S_p^2 is defined as

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

It is easy to show that T has a t_{n+m-2} distribution.

Example: Cedar-apple rust data (cont'd):

```
> variety2  <- c(27,28,57,66,77,49,62)
```

```
> t.test(variety1,variety2,var.equal=T)

        Two Sample t-test
data:  variety1 and variety2
t = -0.501, df = 13, p-value = 0.6247
alternative hypothesis:
      true difference in means is not equal to 0
95 percent confidence interval:
 -29.40555  18.33412

sample estimates:
mean of x mean of y
 46.75000  52.28571
```

Example: Cedar-apple rust data (cont'd):

```
> var.test(variety1,variety2)

        F test to compare two variances

data:  variety1 and variety2
F = 1.499, num df = 7, denom df = 6,
p-value = 0.638
alternative hypothesis:
         true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2631926 7.6728062
sample estimates:
ratio of variances
          1.499006
```

When the variances are not equal, an approximate test is *Welch's modified two-sample t test*, given by

$$T = \frac{(\bar{X} - \bar{Y}) - \Delta_0)}{S_w}$$

where

$$S_w^2 = \frac{S_1^2}{n} + \frac{S_2^2}{m}$$

T has an approximate t_ν distribution, with

$$\nu \approx \left[\frac{c^2}{n-1} + \frac{(1-c)^2}{m-1}\right]^{-1}$$

where

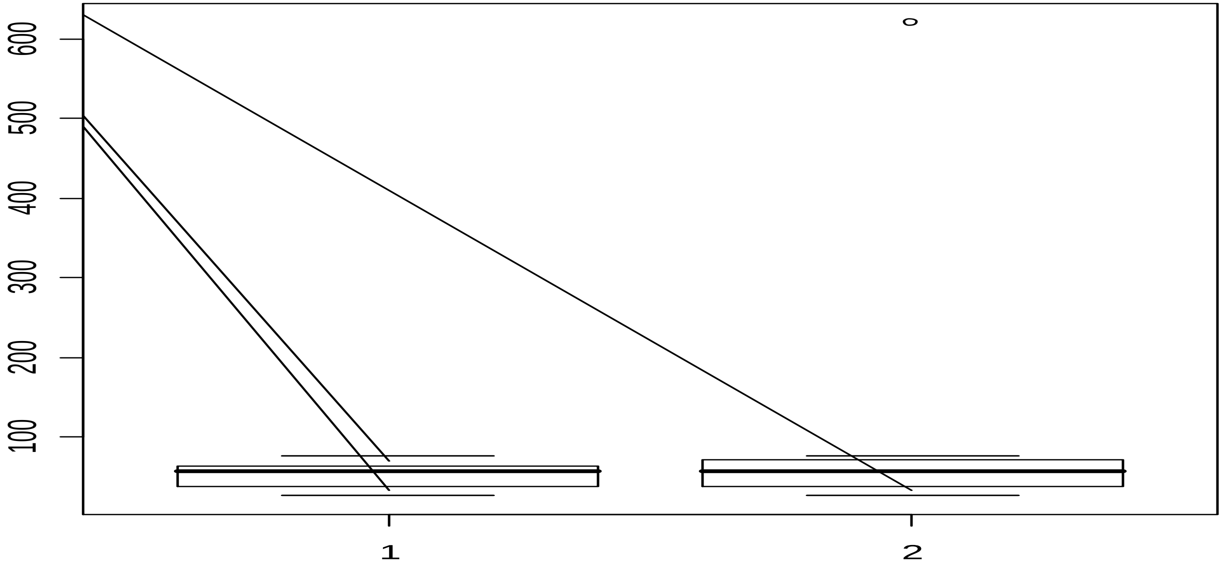
$$c = \frac{S_1^2}{nS_w^2}$$

Example: Cedar-apple rust data (cont'd):

```
> t.test(variety1,variety2)

        Welch Two Sample t-test

data:  variety1 and variety2
t = -0.5082, df = 12.956,
 p-value = 0.6198
alternative hypothesis: true difference in means is not equal
    to 0
95 percent confidence interval:
 -29.07417  18.00275
sample estimates:
mean of x mean of y
 46.75000  52.28571
```

Example: Cedar-apple rust data (cont'd):

```
> variety2  <- c(27,28,57,66,77,49,62)

>variety3  <- c(27,28,57,66,77,49,6200)

> boxplot(variety2,variety3)
```



Example: Cedar-apple rust data (cont'd):

```
t.test(variety1,variety2,var.equal=T)

        Two Sample t-test
data:  variety1 and variety2
t = -0.501, df = 13,
p-value = 0.6247
alternative hypothesis:
      true difference in means is not
      equal to 0
95 percent confidence interval:
 -29.40555  18.33412

sample estimates:
mean of x mean of y
 46.75000  52.28571
```

```
 > variety3  <- c(27,28,57,66,77,49,620)

> t.test(variety1,variety3,var.equal=T)

        Two Sample t-test

data:  variety1 and variety3
t = -1.1151, df = 13, p-value = 0.285
alternative hypothesis: true difference in
     means is not equal to 0
95 percent confidence interval:
 -250.4077   79.9077
sample estimates:
mean of x mean of y
   46.75    132.00
```

Wilcoxon Rank-Sum Test

- Combine the two samples, and rank the observations
- Let T_1 be the sum of the ranks for the observations in the 1st group, and T_2 for the second
- The Wilcoxon rank-sum test is then based on T_W, which has an asymptotic standard normal distribution,

$$T_W = \frac{T_1 - E(T_1 \mid H_o)}{\sqrt{Var(T_1 \mid H_o)}}$$

where

$$E[T_1 \mid H_o] = \frac{n(n+m+1)}{2}$$

and

$$Var(T_1 \mid H_o) = \frac{nm(n+m+1)}{12}$$

In some applications, the Mann-Whitney (U) form of the Wilcoxon rank-sum test is used, due to the relative ease of computing critical values using the latter. The two are related by the equation

$$U_1 = T_1 - \frac{n(n+1)}{2}$$

- The test is most appropriate when the populations have the same shape and differ only in location (e.g., same dispersion) Howeve, the distributions do not have to be symmetric.

Example: Cedar-apple rust data (cont'd):

```
> wilcox.test(variety1,variety2)

        Wilcoxon rank sum test

data:  variety1 and variety2
W = 24, p-value = 0.6943
alternative hypothesis: true mu
is not equal to 0
```

```
> variety3  <-
   c(27,28,57,66,77,49,620)

> wilcox.test(variety1,variety3)

        Wilcoxon rank sum test

data:  variety1 and variety3
W = 22, p-value = 0.5358
alternative hypothesis: true mu
   is not equal to 0
```

Example: Cedar-apple rust data (cont'd):

- Disease that affects apple trees, with symptom, rust-colored spots on apple leaves. Study conducted to study cause.
- In the first year of experiment the number of affected leaves on a random sample of 8 trees from Variety 1
- All red cedar trees within 100 yards of the orchard were then removed and the following year the same Variety 1 trees were examined for affected leaves.

Data:

Year 1		Year 2
Tree	Count	Count
1	38	32
2	10	16
3	84	57
4	36	28
5	50	55
6	35	12
7	73	61
8	48	29

Does the data suggest there is a difference between the mean number of affected leaves in the two periods?

Example: Cedar-apple rust data (cont'd):

Apply 2-sample t-test?

```
> t.test(year1,year2,var.equal=T)

        Two Sample t-test

data:  variety1 and year2
t = 0.9899, df = 14, p-value = 0.3390
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.25070  33.25070
sample estimates:
mean of x mean of y
    46.75     36.25
```

Paired Samples

Consider a study comparing weight gains in rats before and after receiving a new dietary regime. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be the respective weights of n rats. Clearly, there is dependence between obseravtions on the same rat, hence, the two-sample procedures discussed above will not be appropriate.

Let $D_i = X_i - Y_i$, $i = 1, \cdots, n$. Then the problem may be handled based on the D_i's using the one-sample t-test (paired t-test) or the Wilcoxon signed-rank test.

Example: Cedar-apple rust data (cont'd):

```
> t.test(year1,year2,paired=T)

        Paired t-test

data:  variety1 and year2
t = 2.4342, df = 7,
p-value = 0.04514
alternative hypothesis: true difference in
means is not equal to 0
95 percent confidence interval:
  0.2999574 20.7000426
sample estimates:
mean of the differences
                   10.5
```

```
> wilcox.test(year1,year2,paired=T)

        Wilcoxon signed rank test with
continuity correction

data:  year1 and year2
V = 32.5, p-value = 0.04967
alternative hypothesis: true mu is not
equal to 0
```

Measures of Associations

Given a random sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ on n individuals, we are often interested in the degree of association or relationship between X and Y.

1. *Pearson's product-moment correlation coefficient*

The population correlation coefficient is defined as

$$\rho = \frac{E(X - \mu_x)(Y - \mu_y)}{\sigma_x \sigma_y}$$

and has the property $-1 \leq \rho \leq 1$.

The corresponding sample quantity is given by

$$r = \frac{\Sigma_i (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1) S_x S_y}$$

with similar properties as ρ.

For the problem of testing $H_0 : \rho = 0$ vs. $H_1 : \rho > 0$, a suitable test statistic is given by

$$T = \sqrt{\frac{n-2}{1-r^2}}$$

which has a t_{n-2} null distribution.

The test of the more general hypothesis $H_0 : \rho = \rho_0$ vs. $H_1 : \rho > \rho_0$ involves Fisher's Z-transformation. Let

$$Q(r) = \frac{1}{2} ln \left(\frac{1+r}{1-r} \right)$$

Then a test statistic is given by

$$Z = \frac{Q(r) - Q(\rho)}{\sqrt{\frac{1}{n-3}}}$$

which has an apprximate standard normal null distribution.

For purposes of confidence interval construction, one often uses the following inverse hyperbolic tangent transformation. Let $H(r) = atanh(r)$. Then $H(r) - H(\rho)$ has an approximate $N(0, 1/(n-3))$ distribution. A

$100(1-\alpha)\%$ confidence interval for ρ is

$$tanh\left[(H(r) \pm Z_{\alpha/2}\frac{1}{\sqrt{n-3}}\right].$$

Remarks

- The product moment correlation coefficient may not be appropriate for ordinal data
- The sample correlation coefficient r is extremely sensitive to outliers

2. *Spearman's rank correlation coefficient*

Spearman's rho is a measure of association based on ranks.

Given $(X_1, Y_1), \cdots, (X_n, Y_n)$, let d_i denote the difference between the ranks of X_i and Y_i, when the X's and Y's are ranked separately among themselves.

For tied observation, the mean of the ranks are taken.

When there are no ties, Spearman's rho r_s coincides with the Pearson product-moment correlation coefficient applied to the ranks, and is given by

$$r_s = 1 - \frac{6 \Sigma_i d_i^2}{n(n^2 - 1)}$$

3. *Kendall's tau*

Kendall's tau is another rank-based measure of association. Let R_i and S_i denote the ranks of X_i and Y_i, respectively. Then

$$\tau = \frac{1}{\binom{n}{2}} \sum_{i<j} sgn(R_i - R_j)(S_i - S_j)$$

where

$$sgn(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

```
> cor.test(variety1,year2)

        Pearson's product-moment
     correlation

data:  variety1 and year2
t = 3.973, df = 6,
      p-value = 0.007342
alternative hypothesis:
  true correlation is not equal to 0
95 percent confidence interval:
 0.3662168 0.9725356

sample estimates:
     cor
0.851221
```

```
> cor.test(variety1,year2,method="sp")

        Spearman's rank correlation rho

data:  variety1 and year2

p-value = 0.002232
alternative hypothesis:
         true rho is not equal to 0

sample estimates:
      rho
0.9285714
```

Goodness-of-Fit Tests

CDF Test

Let $X_1, \cdots, X_n$ be a random sample from F. The empirical distribution function, F_n, is defined as

$$F_n(x) = \frac{\Sigma_i I[X_i \leq x]}{n}$$

for $x \epsilon R$. As an estimator of F, F_n has several desirable properties.

One way of assessing goodness-of-fit is by looking at a plot of F_n and F, superimposed. The R function *cdf.compare()* provides such plots.

Chi-Square Test of Goodness of Fit

Let $X_1, \cdots, X_n$ be a random sample from F. Group the observations $x_1, \cdots, x_n$ into k classes. Let m_i and M_i be the observed and expected, under F, counts in the i'th class, respectively. The chi-square test of goodness-of-fit is given by:

$$\chi^2_G = \sum_i \frac{(m_i - M_i)^2}{M_i}$$

and has an approximate χ^2_{k-1} distribution under F. In R, the function *chisq.gof* may be used.

The Kolmogorov-Smirnov Test

Let $X_1, \cdots, X_n$ be iid F. Then the one-sample K-S test is given by

$$T_{KS} = sup_x \mid F_n(x) - F(x) \mid$$

In the case of two samples, let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ be independent random samples from F and G, respectively. Then the K-S test for comparing the two distributions is given by:

$$T_{KS} = sup_x \mid F_n(x) - G_m(x) \mid$$

In *R*, the procedures are implemented using *ks.gof()*.

Bootstrap Tests

The bootstrap may be used to construct approximate confidence intervals and test critical points, when the associated distributions are not tractable.

Consider the problem of comparing two means μ_1 and μ_2, when sampling $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ from F and G, respectively.

To implement the bootstrap, one draws pseudo-random samples from F_n and G_m under H_0 : $\mu_1 = \mu_2$.

- Draw with replacement, B pseudo-samples: $X_1^*, \cdots, X_n^*$ from F_n and and $Y_1^*, \cdots, Y_m^*$ from $\tilde{G}_m$, where $\tilde{G}_m$ is the edf of $\tilde{Y}_1, \cdots, \tilde{Y}_m$, and

$$\tilde{Y}_i = Y_i + \bar{X} - \bar{Y}$$

For the b'th pseudo-sample, compute

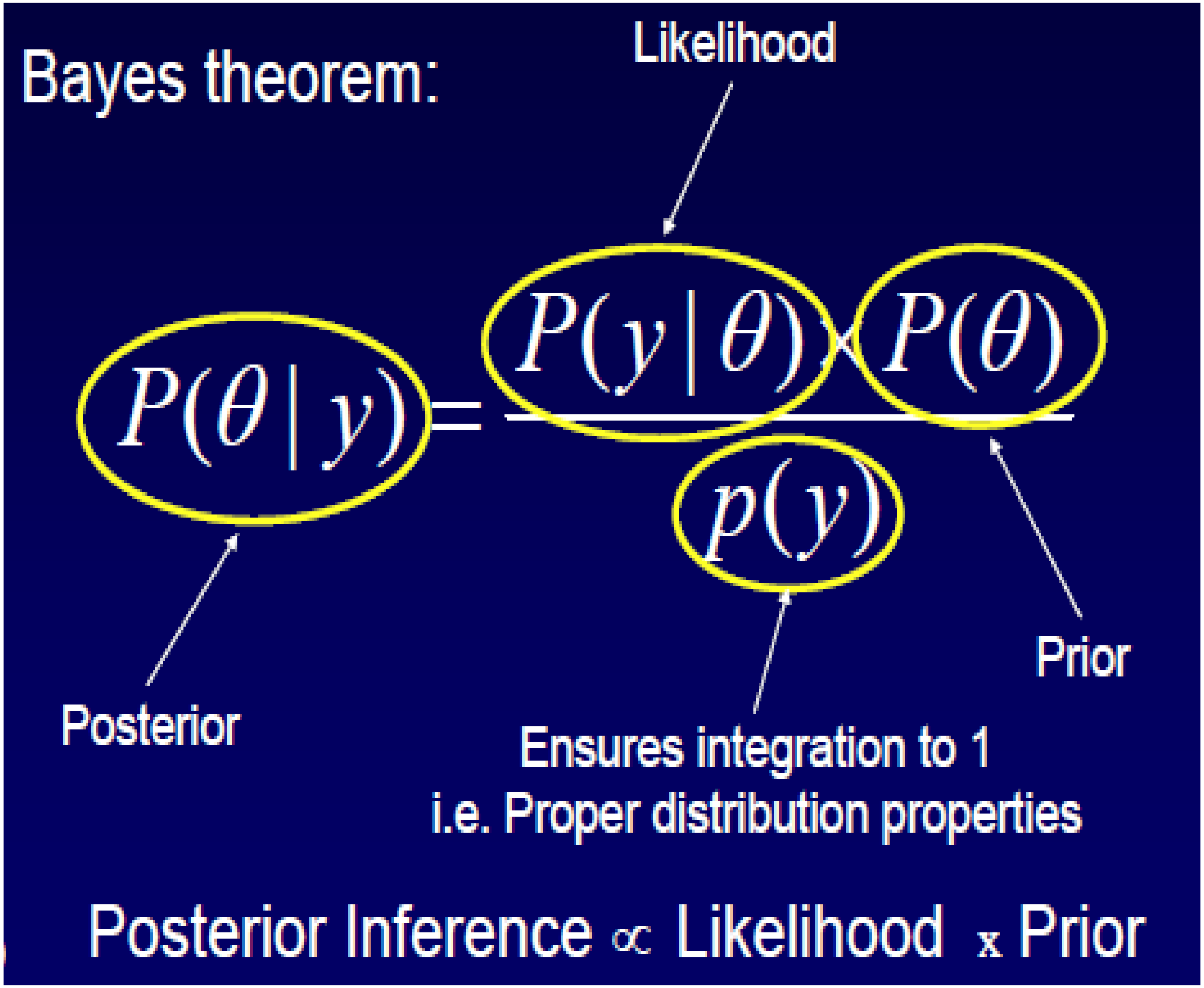
$$Z_b^* = \frac{(\bar{X}^* - \bar{Y}^*)}{\sqrt{\frac{S_1^*}{n} + \frac{S_2^*}{m}}}$$

- Calculate the p-value as

$$p_B = \frac{\Sigma_b^B I[\mid Z_b^* \mid \geq z_{obs}]}{B}$$

where z_{obs} is the observed value of the test statistic based on the original sample.

Bayesian Statistics



Prior Distributions

- Non-informative priors
  - Vague priors, reference priors, flat priors
  - Have minimal impact on the posterior distribution, e.g.,

    if θ is normal(θ_o,$\sigma^2{}_o$), pick large $\sigma^2{}_o$

    Or, if $0 \leq \theta \leq 1$, use U(0,1) as a prior for θ
- Informative priors
  - Based on expert opinion
  - From previous studies

Conjugate Prior Distributions

Family	Conjugate Prior for θ
Bin (n, θ)	$\theta \sim$ Beta (α, β)
$\mathcal{P}(\theta)$	$\theta \sim \mathcal{G}(\alpha, \beta)$
$\mathcal{N}(\theta, \sigma^2)$; σ^2 known	$\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$
$\mathcal{G}(\alpha, \theta)$; α known	$\theta \sim \mathcal{G}(\delta_0, \gamma_0)$
$\mathcal{N}(\alpha, \theta)$; α known	$\theta^{-1} \sim \mathcal{G}(\delta_0, \gamma_0)$

The likelihood, considered in terms
of the parameter(s), has a kernel in the same form as the prior distribution.

Exponential families have conjugate priors in general.

Conjugate priors can be either informative or non-informative.

The Jeffreys' Prior

A general approach: Define the prior based on the Fisher information *I of the* likelihood

$$I(\theta \,|\, \mathrm{x}) = -E_x \left(\frac{\partial^2 \ln \ell(\theta \,|\, \mathrm{x})}{\partial \, \theta^2} \right)$$

So that

$$p(\theta) \propto \sqrt{I(\theta \,|\, \mathrm{x})}$$

	Frequentist	**Bayesian**
Prior information	Informally used in design	Used formally to priors
Interpretation of the parameter of interest	A fixed state of nature	An unknown quantity with probability distribution
Approach	"How likely is the data, given a particular value of the parameter?"	Use posterior distribution of the parameter to calculate and use of the posterior distribution in formal decision analysis
Presentation of results	P-values, confidence intervals	Predictive probabilities

Credible Intervals

a $100(1-\alpha)\%$ *two-sided Bayesian credible interval for a single parameter θ with equal posterior tail probabilities*

$$[\theta_{\alpha/2}, \theta_{1-\alpha/2}]$$

where the two end points are the $\alpha/2$ *and* $1-\alpha/2$ *quantiles of the* posterior distribution of θ

Bayesian vs Frequentist Inference

- Bayesian estimates with flat priors equivalent to ML estimates
- Asymptotically, priors would be irrelevant and Bayesian and frequentist results would converge.
- Frequentist one-tailed tests equivalent to credible intervals. But, Bayesian methods poor for two-tailed tests.
  - E.g., probability of zero in a continuous distribution is zero.
  - In Bayesian paradigm, frequentist p-values violate the likelihood principle.

Bayes Factors

Bayes Factors: Analogues of likelihood ratio tests.

- Ratio of prior and posterior information used to provide evidence in favor of one model specification vs another.

Wish to compare two competing models, M_1 and M_2, with respective parameters parameters, θ_1 and θ_2. and likelihood specifications :

$f_1(x \mid \theta_1)$ and $f_2(x \mid \theta_1)$

$$\text{PosteriorOdds} = \text{Prior Odds/Data} * \text{Bayes Factor}$$

$$\frac{\pi(M_1 \mid x)}{\pi(M_2 \mid x)} = \frac{p(M_1)/p(x)}{p(M_2)/p(x)} \times \frac{\int_{\theta 1} f_1(x \mid \theta_1) p_1(\theta_1) d\theta_1}{\int_{\theta 2} f_2(x \mid \theta_2) p_2(\theta_2) d\theta_2}$$

$$\text{Bayes Factor} = B(x) = \frac{\pi(M_1 \mid x)/\, p(M_1)}{\pi(M_2 \mid x)/\, p(M_2)}$$

NB: Bayes Factors intuitive, but often quite difficult to calculate.

Rule of Thumb

$$\text{Bayes Factor} = B(x) = \frac{\pi(M_1 \mid x) / p(M_1)}{\pi(M_2 \mid x) / p(M_2)}$$

- $B(x) \geq 1$, evidence in favor of model 1
- $1 > B(x) \geq 10^{-1/2,}$, weak evidence against model 1.
- $10^{-1/2} > B(x) \geq 10^{-1}$, substantial evidence against model 1.
- $10^{-1} > B(x) \geq 10^{-2}$, strong evidence against model 1.
- $10^{-2} > B(x)$, decisive evidence against model 1.

Problem Set 2

I) Reading Assignment: Chapter 2,3,4. The Statistical Sleuth: A Course in Methods of Data Analysis. Ramsey & Schafer

II) Consider the RatPupWeight (the weight of rat pups) in R library *nlme:*
data(RatPupWeight, package="nlme")

1)Disregarding the effects of all the other variables, determine whether there is a significant difference between mean weight of rat pubs in the control and active (low/high combined) treatment groups using each of the following procedures:

a) A parametric procedure
b) A non-parametric procedure
c) A re-sampling procedure

2) Discuss the assumption underlying the analyses in (1) above, their validity, and any remedial measures to be taken..
3)Determine whether there is a significant association between weight and Litter Size for each of the following:

a) A parametric procedure
b) A non-parametric procedure

III) Suppose data are drawn from a normal distribution, with mean θ and variance 1. Assume the prior distribution of θ is normal(θ_o,σ^2_o).
1)Determine the posterior distribution of θ, and
2) Indicate the form of a 95% credible interval for the posterior mean of

A useful reference on R may be found at:
http://heather.cs.ucdavis.edu/~matloff/132/NSPpart.pdf (courtesy Jon Fintzi)