Lecture 5: Likelihood Ratio Confidence Intervals & Bayesian Methods for Contingency Tables

Nick Reich / transcribed by Bianca Doone and Guandong Yang

Likelihood Ratio Confidence Intervals

Motivating Example: Poisson Sample-Mean Estimation (1-parameter Poisson Mean)

- The probability mass function (pmf) for the Poisson distribution is defined as:

$$p(y|\mu) = \frac{e^{-\mu}\mu^y}{y!},\ y = 0, 1, ... \text{ (non-negative integers)}$$
$$\text{with } \mathbb{E}(Y) = \mathbb{V}(Y) = \mu$$

- Given observations $y_1, y_2, ..., y_n$, assuming $y_i \overset{iid}{\sim}$ Poisson(μ), the log-likelihood is defined as:

$$L(\mu|\mathbf{y}) = \log\left(\prod_{i=1}^{n} \frac{e^{-\mu}\mu^{y_i}}{y_i!}\right) = \log\left(\frac{e^{-n\mu}\mu^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!}\right)$$
$$= -n\mu + \log(\mu)\sum_{i=1}^{n} y_i + \text{C}$$
$$\text{where C} = \prod_{i=1}^n y_i! \text{ is a constant.}$$

LR CI's - Motivating Example Cont'd

$$L(\mu|\mathbf{y}) \propto -n\mu + \log(\mu)\sum_{i=1}^{n} y_i$$

- Note that taking the first derivative and setting equal to zero provides us the MLE for the data:

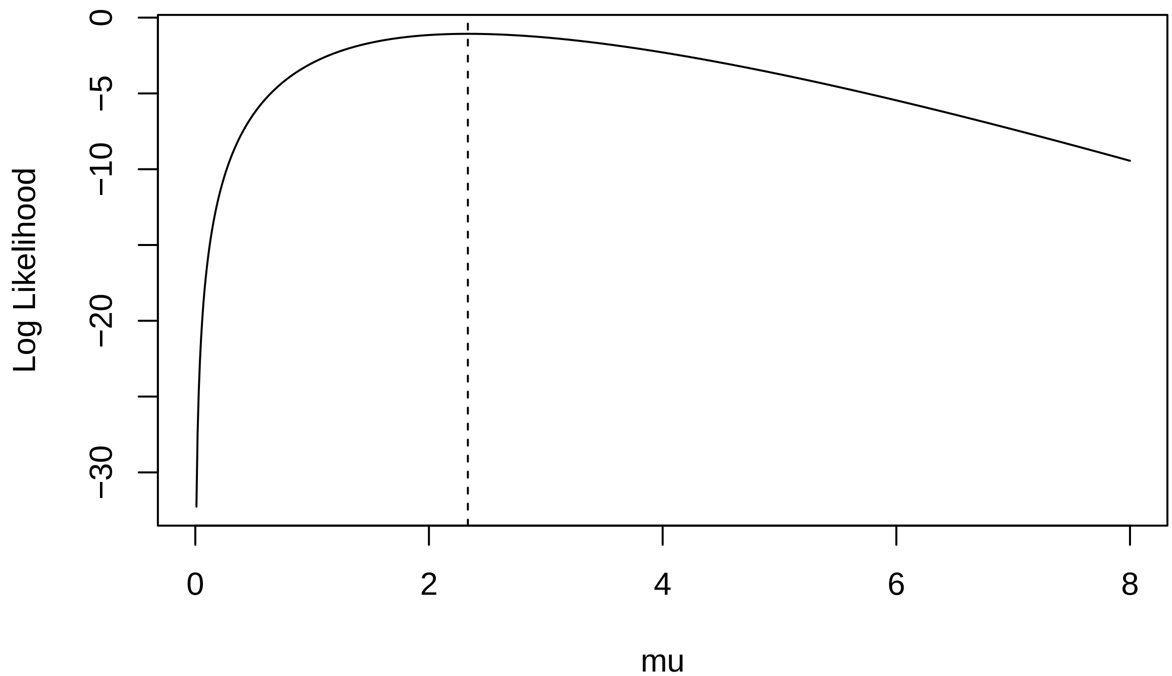
$$\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- Let $\tilde{y} = (2, 2, 3)^T$. Then the log-likelihood is

$$L(\mu|\tilde{y}) \propto -3\mu + 7\log(\mu)$$

maximized at $\hat{\mu} = 7/3$.

LR CI's - Visualization of this likelihood



LR CI's - A Few Notes

- Each point on this curve represents a "fit" to the data.
- In general, adding more data implies that the likelihood is going to be lower.
- Likelihoods always need to be relative (i.e. L_1 vs. L_0).
- Using an absolute scale is not particularly meaningful.

$(1-\alpha)\%$ likelihood-based C.I.

Suppose Θ is a set of parameters, and we let Θ or a subset of Θ vary

Heuristic Definition:

- We are interested in the set Θ for which

$$\text{LRTS}(\Theta) = -2[L(\Theta) - L(\hat{\Theta})] < \chi^2_{df}(1-\alpha)$$

  with $\hat{\Theta}$ as the fixed value at the MLE, and the LR test statistic compared to the $(1-\alpha)^{\text{th}}$ quantile of χ^2_{df}
- The set of Θ where this holds is a $(1-\alpha)\%$ confidence interval with degrees of freedom equal to the number of parameters the likelihood is varying over (free parameters)

LR CI's - Motivating Example Cont'd

Motivating Example: Poisson Sample-Mean Estimation (1-parameter Poisson Mean)

- Let $\Theta = \mu$, then we can define:

$$\text{LRTS}(\mu) = -2[L(\mu) - L(\hat{\mu})]$$

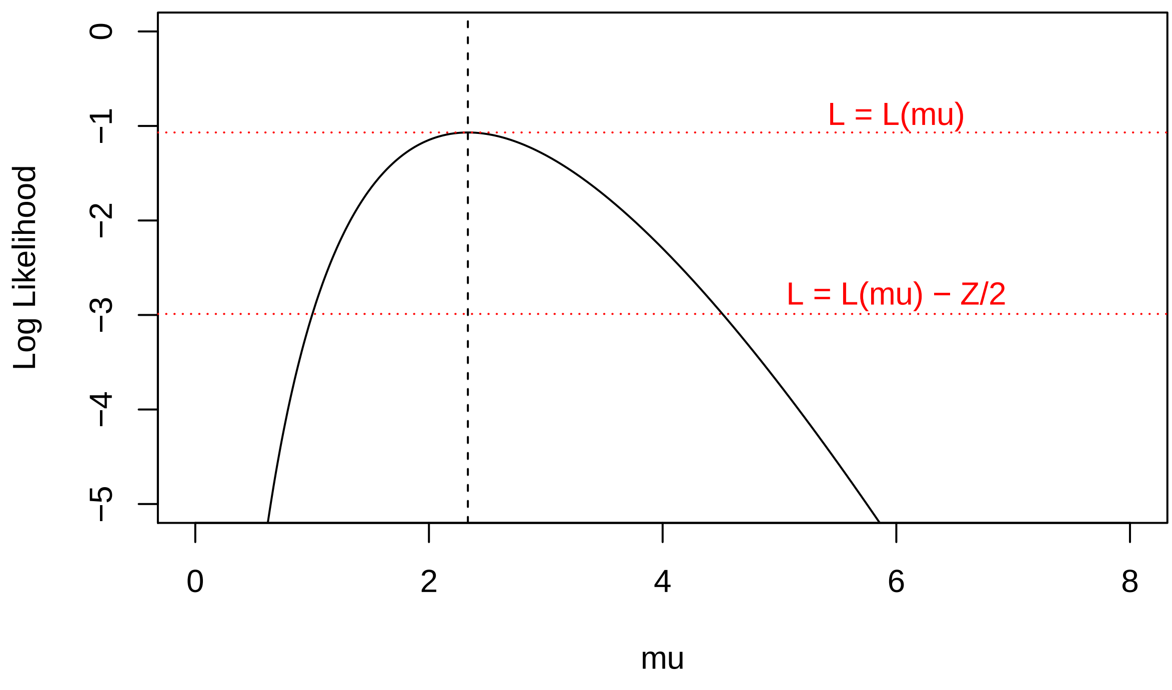
- Setting $\alpha = 0.05$, we can define a 95% confidence interval for μ as

$$\left\{\mu : \text{LRTS}(\mu) < \chi_1^2(.95)\right\}$$
$$= \left\{\mu : -2[L(\mu) - L(\hat{\mu})] < \chi_1^2(.95)\right\}$$
$$= \left\{\mu : L(\mu) > L(\hat{\mu}) - \frac{\chi_1^2(.95)}{2}\right\}$$

LR CI's - Motivating Example Cont'd

Using the MLE, $\hat{\mu} = 7/3$, and the value of the $\chi_1^2(.95) = 3.84$, we derive an approximate 95% confidence interval for μ:

$$\left\{\mu : L(\mu) > L(7/3) - \frac{3.84}{2}\right\} \approx \{\mu : 1.1 \le \mu \le 4.5\}$$



Bayesian Method for Contingency Tables

Bayesian methods for contingency tables may be a good alternative to small sample-size methods because there is less reliance on large sample theory...but could be sensitive to ***prior choice****.*

- Supplement on impact of priors: See Chapter 3.6.4 in Agresti.

Beta/Binomial Example - Handedness

Q: Are women and men left-handed at the same rate?

Gender	RH	LH	Total
Men	43	9	$n_1 = 52$
Women	44	4	$n_2 = 48$
Total	77	13	100

In other words:

- Is there a difference in the proportions of men who are left-handed and women who are left-handed?

$$H_0 : \Pr(\text{left-handed}|\text{male}) = \Pr(\text{left-handed}|\text{female})$$

- Is the difference between left-handed men and left-handed women equal to zero?

$$H_0 : \Pr(\text{left-handed}|\text{male}) - \Pr(\text{left-handed}|\text{female}) = 0$$

Beta/Binomial Example - Handedness

When we have 2×2 table, the chi-square test for independence is equal to two-sided test for different proportion.

```
dat <- matrix(c(43, 9 ,44, 4), ncol =2, byrow = T)
chi <- chisq.test(dat, correct = F)
dif <- prop.test(dat, correct = F)
```

Tests	Test Statistics	DF	P-vlaue
Chi-square test	1.777	1	0.182
Difference Two proportion	1.777	1	0.182

Beta/Binomial Example - Handedness

- **Probability Structure**:
  - Men who are left-handed: $Y_1 \sim \text{Bin}(n_1, \pi_1)$
  - Women who are left-handed: $Y_2 \sim \text{Bin}(n_2, \pi_2)$
- **Observed Data**:
  - $(y_1, y_2) = (9, 4)$
  - $(n_1, n_2) = (52, 48)$
- Let us assign a Uniform prior onto π_1 and π_2, such that
  - $\pi_1 \sim U(0, 1)$ (also considered $\sim \text{Beta}(1, 1)$)
  - $\pi_2 \sim U(0, 1)$
- Because the Beta distribution is a *conjugate prior* to the Binomial likelihood, the **posterior** distribution for pi_1 and π_2 is

$$p(\pi_1|y, n) \sim \text{Beta}(y_1 + 1, n_1 - y_1 + 1)$$

$$p(\pi_2|y, n) \sim \text{Beta}(y_2 + 1, n_2 - y_2 + 1)$$

Bayesian Method - Computational Technique

1. Simulate N independent draws from $p(\pi_1|y, n)$ and $p(\pi_2|y, n)$
2. Compute $\theta_i, i = 1, ..., N$
3. Plot empirical posterior
4. Calculate summary statistics

Bayesian Method - Multinomial/Dirichlet

- Suppose y is a vector of counts with number of observations for each possible outcome, j
- Then, the likelihood can be written as

$$p(y|\theta) \propto \prod_{j=1}^{k} \theta_j^{y_i}$$

  where $\sum_j \theta_j = 1$ and θ is a vector of probabilities for j.
- The conjugate prior distribution is a multivariate generalization of the Beta distribution: **The Dirichlet**

Bayesian Method - Multinomial/Dirichlet

- We set the Dirichlet distribution as the prior for θ: $\theta \sim \text{Dir}(\alpha)$, with pdf:

$$p(\theta|\alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}$$

  where α is a hyper parameter, and $\theta_j > 0, \sum_j \theta_j = 1$
- The posterior distribution can then be derived as

$$p(\theta|y) \sim \text{Dir}\begin{pmatrix} \alpha_1 + y_1 \\ \alpha_2 + y_2 \\ \vdots \\ \alpha_k + y_k \end{pmatrix}$$

- Plausible "non-informative" priors
  - Set $\alpha_j = 1, \forall j$ gives equal density to any vector θ such that $\sum_j \theta_j = 1$
  - Set $\alpha_j = 0, \forall j$ (improper prior) gives a uniform distribution in $\log(\theta_j)$ (if $y_i > 0, \forall j$, we have a proper posterior)

Multinomial/Dirichlet - Example in R

Adapted from Bayesian Data Analysis 3

- A poll was conducted with $n = 1447$ participants, with the following results:
  - Obama: $y_1 = 727$
  - Romney: $y_2 = 583$
  - Other: $y_3 = 137$

- The estimand of interest is $\theta_1 - \theta_2$
- Assuming simple random sampling, we have

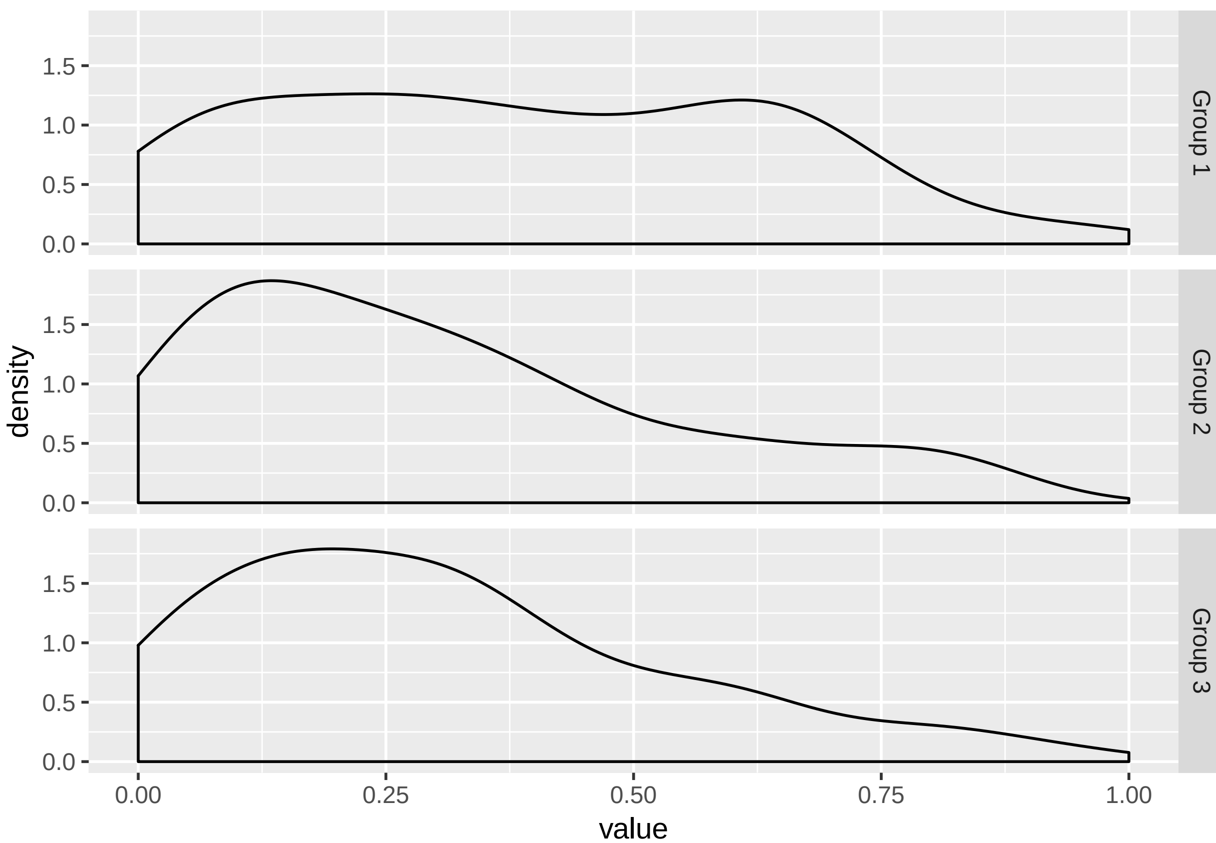
$$(y_1, y_2, y_3) \sim \text{Multinomial}(n, \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix})$$

- We now apply the same computational technique as in the univariate case...

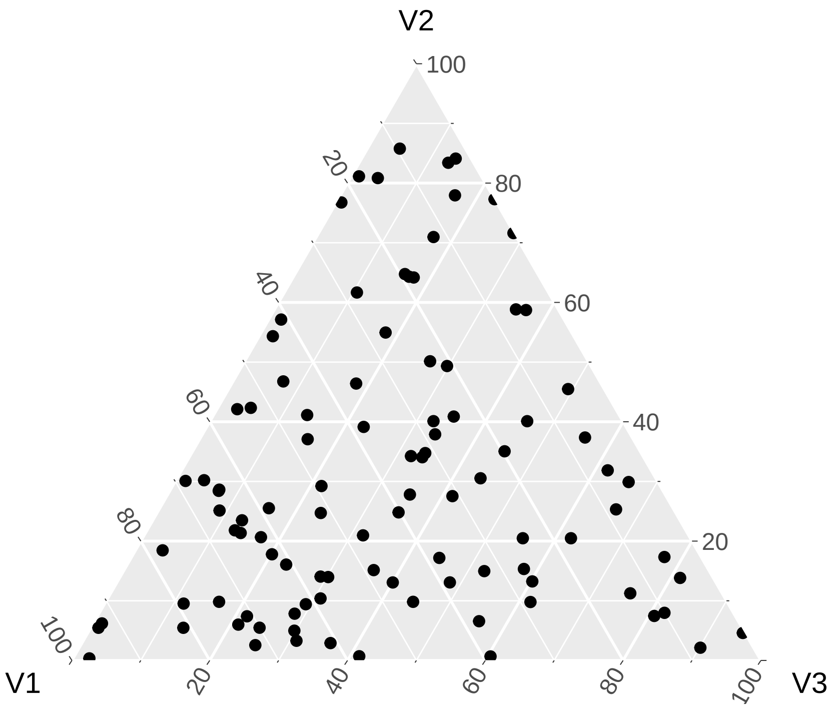
Multinomial/Dirichlet - Example in R

```
#data
y <- c(727, 583, 137)

#"uniform" hyperparameter
a <- c(1,1,1)

#prior
pri <- rdirichlet(100, a)

#Generate Posterior
postr <- rdirichlet(1000, y+a)
```

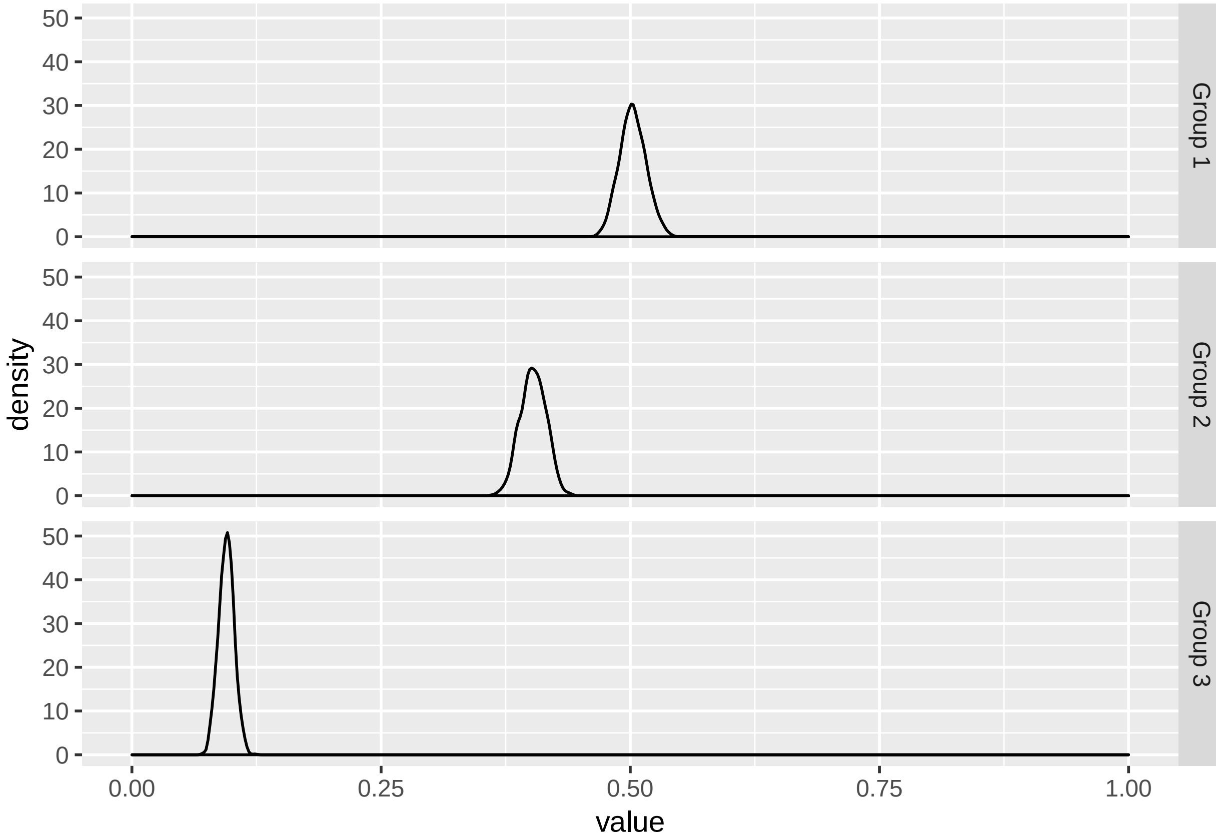
Multinomial/Dirichlet - Visualization of Prior



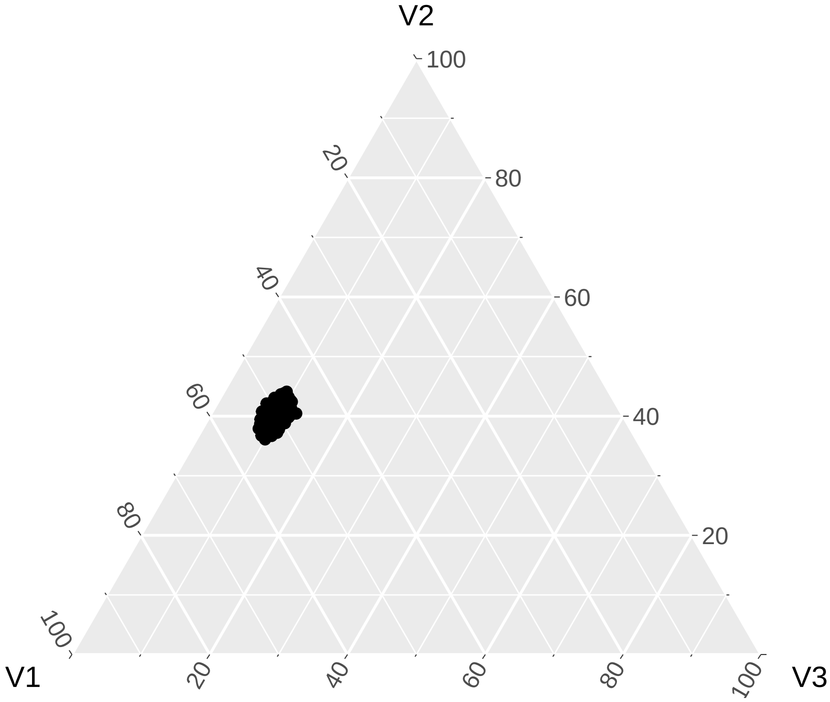
Multinomial/Dirichlet - Visualization of Prior (3D)



Multinomial/Dirichlet - Visualization of Posterior

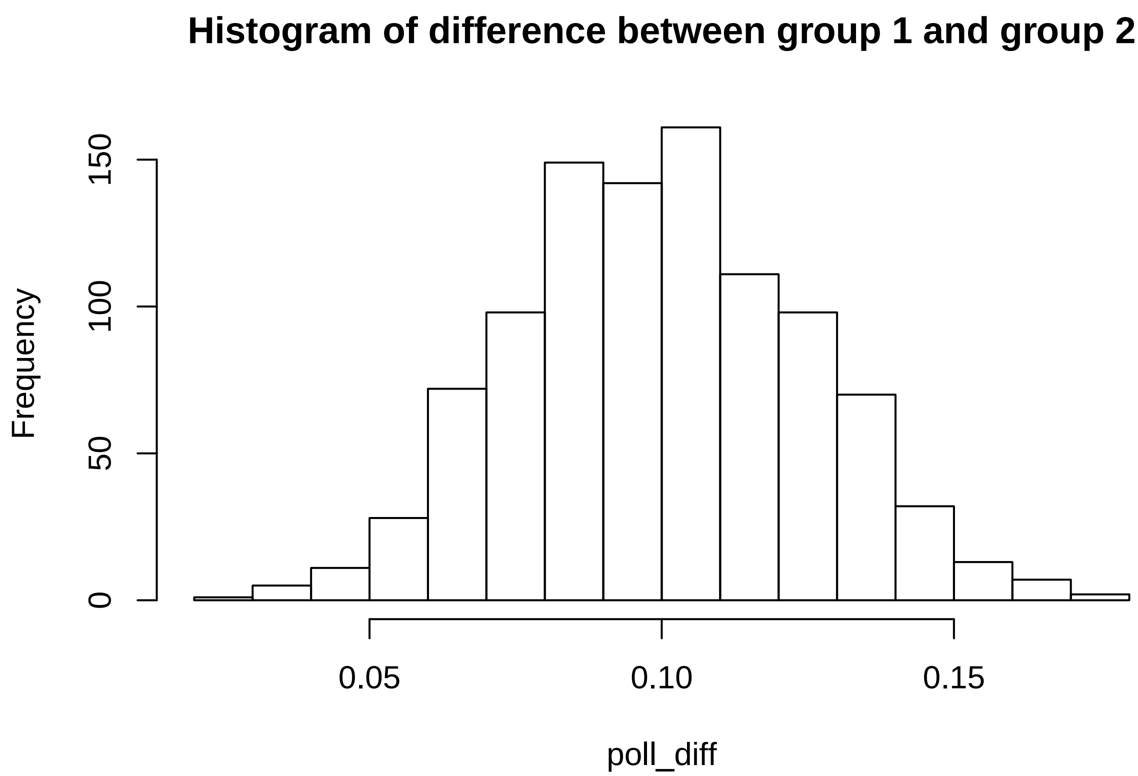


Multinomial/Dirichlet - Visualization of Posterior (3D)



Multinomial/Dirichlet - Summary Statistics

```
poll_diff <- postr[,1]-postr[,2]
hist(poll_diff, main = main)
```



Multinomial/Dirichlet - Summary Statistics

```
### Point Estimates
mean(poll_diff)
```

```
## [1] 0.09999157
```

```
### P-value
mean(poll_diff >0)
```

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

```
### 95% CI
quantile(poll_diff, c(.025, .975))
```

```
##       2.5%      97.5%
## 0.05538754 0.14811648
```