Lecture #14

- Recall binomial r.v. count the number of heads on n flips of a coin with p the probability of heads on a flip, under the assumption the flips are iid.
- So, we write $X \sim Bin(n,p)$ to describe the above, where n and p are parameters.

ex: Suppose we flip a coin 1000 times and observe 575 heads. Evaluate the hypothesis test

$$H_0: p = 1/2$$
$$H_1: p \neq \frac{1}{2}$$

at the $\alpha = .05$ level.

Well, we have $X \sim Bin(p, 1000)$ with p unknown. We observed $X = 575$, and the MLE for p is thus $\frac{X}{n} = .575 = \hat{p}$. By asymptotic normality of $\hat{p}$, we may run a Wald test with

$$W = \frac{\hat{p} - p_0}{\hat{se}}, \quad \text{where} \quad \hat{p} = .575$$
$$p_0 = .5$$
$$\hat{se} = \sqrt{\frac{\hat{p}(1-\hat{p})}{1000}} \approx .0156$$

So, we check: $W > z_{\alpha/2}$

$$\Leftrightarrow \frac{.575 - .5}{.0156} > 1.96$$

$$\Leftrightarrow 4.7977 > 1.96 \quad \checkmark$$

So we reject H_0.

We can also construct a $(1-\alpha)$ confidence interval as

$$\left(\hat{p} - Z_{\alpha/2} \cdot \hat{se},\ \hat{p} + Z_{\alpha/2} \cdot \hat{se}\right) = (.5444, .6056)$$

Since $p_0 \notin (.5444, .6056)$, we fail to retain H_0 (hypothesis tests and confidence intervals can both be used to reject H_0).

/

· What if we don't have two possible outcomes, as in a Binomial r.v., but rather has K possible outcomes, where the i^{th} outcome ours with probability p_i?

Defn: A multinomial r.v. $(X_1, \ldots, X_K)$ with parameters $(n, \vec{p})$, $\vec{p} = (p_1, \ldots, p_K)$ is a K-dimensional random vector with probability distribution

$$\sum_{i=1}^{n} p_i = 1$$

$$\mathbb{P}\left(X_1 = x_1, \ldots, X_K = x_K ; n, \vec{p}\right) = \begin{cases} \dfrac{n!}{\prod_{i=1}^{K} x_i!} \prod_{i=1}^{K} p_i^{x_i}, & \sum_{i=1}^{K} x_i = n \text{ and } \{x_i\}_{i=1}^{K} \text{ are non-negative integers} \\ 0, & \text{else} \end{cases}$$

· Basically, we flip/roll a K-sided die n times, where the i^{th} side occurs with probability p_i. The multinomial r.v. is a vector of length K, where i^{th} entry counts the number of times the i^{th} side is observed.

- Suppose we have a "fair" K-sided die. Then the number of observations of each side should be equal (in expectation), and we have $\vec{p} = (p_1, \dots, p_K)$

$$= \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$$

How could one evaluate a claim that the die is fair (or any claim about the parameter vector $\vec{p}$)?

- Consider the likelihood function associated to an observation $(x_1, \dots, x_K)$ and parameter vector $\vec{p} = (p_1, \dots, p_K)$:

$$\mathcal{L}(\vec{p}) = \mathcal{L}(p_1, \dots, p_K) = \frac{n!}{\prod_{i=1}^{K} x_i!} \cdot \prod_{i=1}^{K} p_i^{x_i}$$

$$\Rightarrow \log\left(\mathcal{L}(p_1, \dots, p_K)\right) = \log\left(\frac{n!}{\prod_{i=1}^{K} x_i!}\right) + \sum_{i=1}^{n} x_i \log(p_i)$$

Now, we differentiate with respect to p_i:

$$\frac{\partial}{\partial p_i}\left(\log\left(\mathcal{L}(p_1, \dots, p_K)\right)\right) = \frac{x_i}{p_i} = 0 \ ???$$

This is strange... we cannot seem to solve the MLE problem by differentiating as before. Why?

We forgot the constraint that $\sum_{i=1}^{k} p_i = 1$! We really have a constrained optimization problem:

$$\max \quad \log\left(\mathcal{L}(p_1,\dots,p_k)\right) = \log\left(\frac{n!}{\prod_{i=1}^{k} x_i!}\right) + \sum_{i=1}^{k} x_i \log(p_i)$$

$$\text{subject to} \quad \sum_{i=1}^{k} p_i = 1$$

$\boxed{Q}$: How to solve such a constrained problem?

$\boxed{A}$: Lagrange multipliers!

constraint $\sum_{i=1}^{k} p_i = 1$

$\Leftrightarrow 1 - \sum_{i=1}^{k} p_i = 0$

$$\underbrace{F(p_1,\dots,p_k,\lambda)}_{\text{"Lagrangian"}} = \log\left(\mathcal{L}(p_1,\dots,p_k)\right) + \lambda\left[1 - \sum_{i=1}^{k} p_i\right]$$

Now differentiating in p_i does the trick: $\frac{\partial}{\partial p_i} F(p_1,\dots,p_k,\lambda) = 0$

$$\Leftrightarrow \quad \frac{x_i}{p_i} - \lambda = 0$$

$$\Leftrightarrow \quad p_i = \frac{x_i}{\lambda}$$

Finally, since $\underset{\overset{\shortparallel}{1}}{\sum_{i=1}^{k} p_i} = \sum_{i=1}^{k} \frac{x_i}{\lambda} = \frac{1}{\lambda}\sum_{i=1}^{k} x_i = \frac{1}{\lambda}\cdot n$, we conclude $\lambda = n$

and hence the MLE for $\vec{p} = (p_1, \ldots, p_k)$ is $\vec{p}_{MLE} = (\hat{p}_1, \ldots, \hat{p}_k)$

$$= \left(\frac{X_1}{n}, \ldots, \frac{X_k}{n} \right).$$

Natural enough.

· So, we have a good (optimal?) point estimator for $\vec{p} = (p_1, \ldots, p_k)$. To do a hypothesis test, we need some notion of variance / standard error. Next time!