1 Probability Review

Multinomial Distribution

$$\Pr(X_1 = x_1, \dots, X_r = x_r) = \binom{n}{x_1, \dots, x_r} \prod_{i=1}^{r} p_i^{x_i}$$

Mean Square Error (MSE)

$$\mathsf{E}\{(Y-c)^2\} = \mathsf{var}(Y) + \{\mathsf{E}(Y) - c\}^2$$

$$\mathsf{E}\{(Y-c)^2|x\} = \mathsf{var}[Y|x] + \{\mathsf{E}[Y|x] - c\}^2$$

which are special cases of $\mathsf{E}(Y^2) = \mathsf{var}(Y) + [\mathsf{E}(Y)]^2$. MSE is minimized if and only if $c = \mathsf{E}(Y)$ or $\mathsf{E}[Y|x]$.
Usually the formula for $\mathsf{E}[Y|x] = f(x)$ is determined from observations/data and x can be a vector of realisations from covariates.

$$\mathsf{MSE}_{\mathsf{empirical}} = \frac{1}{n}\sum_{i=1}^{n}\{\mathsf{E}[Y|x_i] - y_i\}^2$$

In the real world, we have different realisations x_i of the random variable X, hence the mean MSE is

$$\frac{1}{n}\sum_{i=1}^{n}\mathsf{var}[Y|x_i] \approx \mathsf{E}(\mathsf{var}[Y|X]) \le \mathsf{var}(Y)$$

Analysis of Variance (ANOVA)
involves breaking of variance into components

$$\mathsf{var}(Y) = \mathsf{E}(\mathsf{var}[Y|X]) + \mathsf{var}(\mathsf{E}[Y|X])$$

1.1 Distributions

χ^2_1 distribution
Let $Z \sim \mathcal{N}(0,1)$. $V = Z^2$ has a χ^2 distribution with 1 degree of freedom

$$f(v) = \frac{1}{\sqrt{2\pi}} v^{-1/2} e^{-v/2}$$

Gamma distribution

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, t \ge 0$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

χ^2_n distribution
Let $V_1, \dots, V_n$ be IID χ^2_1

$$V = \sum_{i=1}^{n} V_i$$

has a χ^2_n distribution with n degrees of freedom

t distribution
Let $Z \sim \mathcal{N}(0,1)$ and $V \sim \chi^2_n$ be independent

$$t_n = \frac{Z}{\sqrt{V/n}}$$

has a t distribution with n degrees of freedom

F distribution
Let $V \sim \chi^2_m$ and $W \sim \chi^2_n$ be independent

$$F_{m,n} = \frac{V/m}{W/n}$$

has an F distribution with (m,n) degrees of freedom
*Note: $t_n^2 = F_{1,n}$

1.2 Sample Variance

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

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

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

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

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

2 Survey and Random Sampling

Let $X_1, \dots, X_N$ be random draws without replacement from a population of size N with mean μ and variance σ^2.

$$\mathsf{cov}(X_i, X_j) = -\frac{\sigma^2}{N-1} \forall i \ne j$$

$$\mathsf{var}(\bar{X}) = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$$

2.1 Exchangeable

RV's $Y_1, \dots, Y_k$ are exchangeable if all reordered vectors have the same distribution as $(Y_1, \dots Y_k)$. i.e. for any permutation π on $\{1, \dots, K\}$,

$$(Y_{\pi(1)}, \dots, Y_{\pi(k)}) \stackrel{d}{=} (Y_1, \dots, Y_k)$$

2.2 Estimate and Estimator

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- μ, σ, σ^2 are **parameters**
- $\bar{x}$ is an **estimate** of μ
- $\bar{x}$ is a realisation of the **estimator** $\bar{X}$
- **Standard Error (SE)** of the estimate (a number) is defined as the SD of the estimator

  $$\mathsf{SE} = \mathsf{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

  which is how much $\bar{X}$ fluctuates around μ (a number) estimated from the data
- Estimate of σ

- Biased estimate of σ^2

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

- Unbiased estimate of σ^2 (preferred)

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$E(s^2) = \sigma^2$$

How to estimate μ?

- μ is estimated by $\bar{x}$
- Error in $\bar{x}$ is measured by the SE:

$$\text{SD}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

which is **estimated** by $\frac{s}{\sqrt{n}}$ since σ is unknown
- **Conclusion**: μ is estimated as $\bar{X}$, give or take $\frac{s}{\sqrt{n}}$

$$\text{SE estimated by } \frac{s}{\sqrt{n}} = \frac{\sqrt{\frac{n}{n-1}} \times \text{SD}}{\sqrt{n}}$$

$$\text{where SD } = \hat{\sigma}$$

How to estimate p?

- $\hat{p}$ is the estimator of p

$$E(\hat{p}) = p$$

$$\text{var}(\hat{p}) = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$$

$$\text{SE} = \text{SD}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

which is **estimated** by realisations of $\hat{p}$

2.3 Interval estimation

2.3.1 Definitions

- For sufficiently large n,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

- The p-quantile of $Z \sim \mathcal{N}(0,1)$ is the number q such that

$$\Phi(q) = \Pr(Z \le q) = p$$
$$q = \Phi^{-1}(p)$$

```
q <- qnorm(p)
p <- pnorm(q)
```

- For $0 < p < 0.5$, let z_p be such that

$$\Pr(Z > z_p) = p$$

$$z_p = \Phi^{-1}(1-p)$$

In other words, $z_p = (1-p)$-quantile of Z

2.3.2 CI Estimation

- For large n,

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

$$\Pr\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\frac{\alpha}{2}}\right) \approx 1-\alpha$$

$$\Pr\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha$$

where the above, $\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$ is a random interval. Realisation $\bar{x}$ of $\bar{X}$ gives the realised interval
- $(1-\alpha)$-CI for μ is of the form

$$\left(\text{estimate} - z_{\frac{\alpha}{2}}\text{SE}, \text{estimate} + z_{\frac{\alpha}{2}}\text{SE}\right)$$

2.3.3 Exact CI

- Let $t_{\frac{\alpha}{2},n-1}$ be the number such that

$$\Pr(t_{n-1} > t_{\frac{\alpha}{2},n-1}) = \alpha/2$$

- **[Important]** Exact CI only works if $X \sim \mathcal{N}(\mu, \sigma^2)$ and x_i's are realisations from IID Normal Distribution
  * CI is exact means that $\Pr(\mu$ is within the interval$)$ is exactly $1-\alpha$
- $(1-\alpha)$-CI for μ is

$$\left(\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right)$$

2.4 Bias in Survey

Famous example: US presidential election survey conducted by *Literary Digest* in 1936

2.4.1 Bias in Measurement

- $x_1, \dots, x_n$ are realisations of random draws $X_i, \dots, X_n$ from a population with mean $\mu + b$ and variance σ^2
- SE $= \sigma/\sqrt{n}$ measures how far $\bar{x}$ is from $E(\bar{X}) = \mu + b$
- **Definition of Bias**

$$\text{Bias of estimate} = \text{E(estimator)} - \text{parameter}$$

- MSE

$$\text{E}(\bar{X} - \mu)^2 = \text{var}(\bar{X}) + \{\text{E}(\bar{X}) - \mu\}^2$$
$$\text{MSE} = \text{SE}^2 + \text{bias}^2$$

However μ is unknowable, hence it is not possible to remove bias unless we make very careful observations

3 Parameter Estimation

Assuming data $x_1, \dots, x_n$ are realisations of IID RV's $X_1, \dots, X_n$ with density $f(x|\theta)$, estimate θ.
The parameter θ lies in $\Theta \subseteq \mathbb{R}$ where Θ is the parameter space
How to estimate θ from realisations $x_1, \dots, x_n$?

1. Method of moments
2. Method of maximum likelihood

3.1 Method of moments

Let $\hat{\theta}$ be an estimator for θ.
The k-th moments of an RV X is

$$\mu_k = E(X^k)$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k$$

is a realisation of $\hat{\mu_k}$ and is used as estimate for μ_k

$$\hat{\theta} = g(\hat{\mu_1}, \ldots, \hat{\mu_q})$$

is an estimate for θ *e.g.* for Normal RV,

$$g : \begin{bmatrix} x \\ y \end{bmatrix} \rightarrow \begin{bmatrix} x \\ y - x^2 \end{bmatrix}$$

3.2 Monte Carlo Approximation

Needed if formula for θ is complicated/hard to compute the value of its expectation
Rough Steps:

1. Estimate parameters θ using MOM/MLE
2. Generate n realisations $x_1, x_2, \ldots, x_n$ using the estimated parameters and distribution
3. From these n realisations, estimate parameters again, these are realisations of $\hat{\theta}^*$
4. Repeat steps 2 and 3 m times until we get m realisations of parameters θ

$$\mathsf{SE} = \mathsf{SD}(\hat{\theta}) \approx \mathsf{SD}(\hat{\theta}^*)$$

$$\mathsf{Bias} = \mathsf{E}(\hat{\theta}) - \theta \approx \mathsf{E}(\hat{\theta}^*) - \theta_{\mathsf{est.}}$$

5. Finally, θ is around $\theta_{\mathsf{est.}} -$ Bias $\pm$ SE, and the fitted distribution + parameter is called a **statistical model** for the event in question

Note that as $n \to \infty$, $\mathsf{E}(\hat{\theta}^*) \to \theta_{\mathsf{est}} \Rightarrow$ Bias$\to 0$, $\mathsf{E}(\hat{\theta}) \to \theta$.

- Thus, it is **asymptotically unbiased**
- Every MOM estimator is consistent, it goes to the parameter as $n \to \infty$

3.3 Maximum Likelihood Method

Let $x_1, \ldots, x_n$ be realisations of IID RV's $X_1, \ldots, X_n$ with density/mass function $f(x|\theta)$

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} f(x_i|\theta)$$

Find the value of θ that maximises the likelihood

3.3.1 Multinomial Data

$$L(p_1, \ldots, p_r) = p_1^{x_1} \ldots p_r^{x_r} \times c$$

$$l(p_1, \ldots, p_r) = x_1 \log p_1 + \cdots + x_r \log p_r + \log c$$

Since $p_1 + \cdots + p_r = 1$, differentiating l does not work since it's constrained, hence we use the **Lagrangian** function and treating $p_1, \ldots, p_r, \lambda$ as if they are unconstrained

$$\mathcal{L}(p_1, \ldots, p_r, \lambda) = x_1 \log p_1 + \cdots + x_r \log p_r + \lambda(p_1 + \cdots + p_r - 1)$$

3.3.2 Genetics

Chromosomes come in pairs, one from each parent
Locus a subsequence on a chromosome
Alleles different versions of bases at a locus
Genotype an unordered pair of alleles

- Given k different alleles, we can construct $k(k+1)/2$ different genotypes
- Given the genotype proportions, we can calculate the allele proportions
- Given the allele proportions, we can calculate the genotype proportions

Mendel's Laws of inheritance

- The maternal allele is randomly chosen from her two alleles; similarly for the paternal allele
- The two choices are independent

Hardy-Weinberg Equilibrium: A population is in HWE at a locus if the genotype proportions are

$$f(a_i a_j) = \begin{cases} p_i^2 & i = j \\ 2p_i p_j & i \neq j \end{cases}$$

where p_i is the proportion of allele a_i (assumption: random mating, no mutation, no migration)

3.4 Large-Sample Variance of ML Estimator

$$\mathcal{I}(\theta) = -E\left[\frac{d^2 \log f(X)}{d\theta^2}\right]$$

4 Useful Results

4.1 Algebra

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

4.2 Procedures

Framework for statistical inference:

1. Parameter is a simple function of the population, real or hypothetical
2. Data are realisations of IID RV's (if $n \ll N$)
3. Estimate is a realisation of an estimator, whose SD is the SE. For large n, can construct CI.
4. MSE = SE^2 + bias^2

4.3 Multivariable Calculus

- Use Hessian matrix to calculate partial derivatives/-maximum points, and $|H| > 0$