Stats 200: Introduction to Statistical Inference

Samuel Wong
Department of Statistics
Stanford University

Abstract

The course starts with a review of probability, calculus, and random variables. There are definitions for joint distributions, expectation, and variance calculations. Some theorems such as the Law of Large Numbers and the Central Limit Theorem are discussed. Parameter estimation using MoME, MLE, and Bayes are covered. Hypothesis testing and goodness of fit using the Likelihood Ratio tests are then discussed. We move from one sample to two samples, and then to n samples. Both parametric and non-parametric tests are examined, as well as discrete and continuous data.

Contents

1 Chapter 1: Probability

1.1 Calculus Cheat Sheet

1.1.1 Logs

- $log_b(M * N) = log_b M + log_b N$
- $log_b(\frac{M}{N}) = log_b M - log_b N$
- $log_b(M^k) = klog_b M$
- $e^n e^m = e^{n+m}$
- $a^n b^n = (a+b)^n$
- $(b^n)^m = b^{n*m}$

1.1.2 Derivatives

- $(x^n)' = nx^{n-1}$
- $(e^x)' = e^x$
- $(lnx)' = \frac{1}{x}$
- $(fg)' = fg' - f'g$
- $(\frac{f}{g})' = \frac{f'g - g'f}{g^2}$
- $(f(g(x)))' = f'(g(x))g'(x)$

1.1.3 Integration

- $\int x^n dx = \frac{x^{n+1}}{n+1}$
- $\int a^x dx = \frac{a^x}{ln(a)}$
- Integration by Parts: $\int udv = uv - \int vdu$
- Integrating Even Functions (f(-x) = f(x)): $\int_{-a}^{a} f(x)dx = 2\int_0^a f(x)dx$
- Integrating Odd Functions(f(-x) = -f(x)): $\int_{-a}^{a} f(x)dx = 0$

1.2 Probability Measure

1.2.1 Axioms

1. $P(\Omega) = 1$
2. $A \subset \Omega \implies P(A) \geq 0$
3. A_1, A_2 disjoint $\implies P(A_1 \cup A_2) = P(A_1) + P(A_2)$, more generally, if $A_1, ..., A_n$ disjoint, $\implies P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

1.2.2 Properties

1. $P(A^C) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. $A \subset B \implies P(A) \leq P(B)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

1.2.3 Laws

- Commutative Law: $A \cup B = B \cup A$; $A \cap B = B \cap A$
- Associative Law: $(A \cup B) \cup C = A \cup (B \cup C)$; $(A \cap B) \cap C = A \cap (B \cap C)$
- Distributive Law: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$; $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
- Addition Law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Multiplication Law: $P(A \cap B) = P(A|B)P(B)$

1.3 Law of Total Probability

Let $B_1, ..., B_n$ be disjoint with $\bigcup B_i = \Omega$ and $P(B_i) > 0$. Then, $\forall\, i$:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$$

1.4 Bayes' Theorem

Let $B_1, ..., B_n$ be disjoint with $\bigcup B_i = \Omega$ and $P(B_i) > 0$. Then, $\forall\, i$:

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_i P(A \mid B_i)P(B_i)}$$

1.5 Independence

- <u>Pairwise independent</u>: any two are independent
- <u>Mutually independent</u>: all are independent
  $MI \implies PI$

2 Chapter 2: Random Variables

2.1 Mass/Density Functions

- Probability Mass Function (PMF) = Frequency Function = $p(x_i)$ for discrete RV
- Probability Density Function (PDF) = $f(x)$ for continuous RV
  - All values of $f(x) \geq 0$
- Cumulative Distribution Function (CDF) = $F(x) = P(X \leq x)$
  1. $lim_{x \to -\infty} F(x) = 0$
  2. $lim_{x \to \infty} F(x) = 1$
  3. Always non-decreasing

2.2 Independent RVs

$F(x_1, x_2, ..., x_n) = F_{x_1}(x_1)F_{x_2}(x_2)...F_{x_n}(x_n)$

2.3 Discrete Distributions

- Bernoulli
  - $p(x) =$
  $$\begin{cases} p^x(1-p)^{1-x};\ x = 0\ or\ x = 1 \\ 0\ otherwise \end{cases}$$
  - $E(X) = p$
  - $Var(X) = p(1-p)$
- Binomial
  - $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$; $k = 0, 1, ...n$
  - $E(X) = np$
  - $Var(X) = np(1-p)$
- Geometric
  - $p(k) = (1-p)^{k-1}p$; $k = 1, 2, 3...$
  - $E(X) = \frac{1}{p}$
  - $Var(X) = \frac{1-p}{p}$
- Negative Binomial
  - $p(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$
  - $E(X) = \frac{pr}{1-p}$
  - $Var(X) = \frac{pr}{(1-p)^2}$
- Hypergeometric
  - $p(k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$
  - $E(X) = \frac{mr}{n}$
  - $Var(X) = \frac{mr}{n} \frac{n-m}{n} \frac{n-r}{n-1}$
- Poisson
  - $p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$; $k = 0, 1, 2...$
  - $E(X) = \lambda$
  - $Var(X) = \lambda$

2.4 Continuous Distributions

- Uniform
  - $f(x) =$
    $$\begin{cases} \frac{1}{b-a};\ a \leq x \leq b \\ 0\ otherwise \end{cases}$$
  - $E(X) = \frac{a+b}{2}$
  - $Var(X) = \frac{(b-a)^2}{12}$
- Exponential
  - $f(x) = \lambda e^{-\lambda x};\ x \geq 0$
  - $E(X) = \frac{1}{\lambda}$
  - $Var(X) = \frac{1}{\lambda^2}$
- Normal
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}};\ -\infty < x < \infty$
  - $E(X) = \mu$
  - $Var(X) = \sigma^2$
- Gamma
  - $g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t};\ t \geq 0;$ where $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$
  - $E(X) = \frac{\alpha}{\lambda}$
  - $Var(X) = \frac{\alpha}{\lambda^2}$
- Beta
  - $f(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1};\ 0 \leq u \leq 1$
  - $E(X) = \frac{a}{a+b}$
  - $Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$
- Cauchy
  - $f_z(z) = \frac{1}{\pi(z^2+1)};\ -\infty < z < \infty$
  - $E(X) =$ Does not exist
  - $Var(X) =$ Does not exist
  - Cauchy equals Y/X where X, Y independent $N(0,1)$

3 Chapter 3: Joint Distributions

3.1 Theorem: Functional Independence

$X \perp Y \implies g(X) \perp h(Y)$ for any g, h

3.2 Joint Frequency

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\,dy\,dx$$

3.3 Marginal Frequency

$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

3.4 Conditional Frequency

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$\implies f_{XY}(x,y) = f_{Y|X}(y \mid x) \cdot f_X(x)$$

$$\implies f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y \mid x)\, f_X(x)\,dx$$

3.5 Convolution

$p_z(z) = \sum_{x=-\infty}^{\infty} p_X(x) p_Y(z-x)$
$f_z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$

3.6 Multinomial

$$p(x_1, ..., x_r) = \binom{n}{x_1, ..., x_r} p_1^{x_1}\, p_2^{x_2}\, \cdots\, p_r^{x_r}$$

$$\begin{cases} \sum x_i = n \\ \sum p_i = 1 \end{cases}$$

3.7 Bivariate Normal PDF

$$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp(-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}])$$

4 Chapter 4: Expected Variables

4.1 Definitions

4.1.1 Mean

Def: $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

4.1.2 Variance

Def: $Var(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2$

4.1.3 Covariance

Def:

- $\operatorname{Cov}(X, Y) = \operatorname{E}(XY) - \operatorname{E}(X)\operatorname{E}(Y)$

Variance property:

- $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y)$

4.1.4 Correlation coefficient

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

4.1.5 Conditional expectation

$$E(Y \mid X = x) = \begin{cases} \sum_y y\, p_{Y|X}(y \mid x) & \text{if discrete} \\ \int y\, f_{Y|X}(y \mid x)\, dy & \text{if continuous} \end{cases}$$

4.1.6 Laws

$$E(Y) = E(E(Y|X)))$$
$$Var(Y) = Var(E(Y|X)) + E(Var(Y|X))$$

4.1.7 Properties

- $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu, a^2\sigma^2)$
- If X, Y independent RV, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$
- If $Y = aX + b$, $E(Y) = aE(X) + b$
- If $Y = aX + b$, $Var(Y) = b^2 Var(X)$
- $Var(X + Y) = Cov(X + Y, X + Y)$
- If $U = a + \sum_{i=1}^{n} X_i b_i$, $V = c + \sum_{j=1}^{m} d_i Y_j$, then $Cov(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j Cov(X_i, Y_j)$
- If X_i's independent, $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$

4.1.8 Moment generating function

$$M(t) = \begin{cases} \sum_x e^{tx} p(x) & \text{if discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx & \text{if continuous} \end{cases}$$

4.1.9 r^{th} moment

$$\mu_r = \operatorname{E}(X^r)$$

4.2 Theorems

4.2.1 Markov inequality

$P(X \geq t) \leq \frac{\operatorname{E}(X)}{t}$, where X is a RV where $P(X \geq 0) = 1$

4.2.2 Chebyshev inequality

$$P(|X - \mu| > t) \leq \frac{Var(X)}{t^2}$$

$$P(|\overline{X}_n - \mu| > k\sigma) \leq 1/k^2$$

4.2.3 Moment generating function theorems

- $M^{(r)}(0) = \mathrm{E}(X^r)$
- Characteristic Function: $E(e^{itX})$
- $Y = a + bX \implies M_Y = e^{at} M_X(bt)$
- $Z = X + Y,\ X \perp Y \implies M_Z = M_Y M_X = E(e^{tX})E(e^{tY})$

4.3 Common MGF Derivations

- Geometric MGF
  $E(e^{tk}) = \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} = \frac{p}{1-p} \sum_{k=1}^{\infty} [e^t(1-p)]^k$; using the property $\sum_{k=1}^{\infty} q^k = \frac{q}{1-q}$; $\Rightarrow \frac{p}{1-p}[\frac{e^t(1-p)}{1-e^t(1-p)}] = \frac{pe^t}{1-e^t(1-p)}$
- Poisson MGF
  $E(e^{tk}) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{e^{tk}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e)^{tk}}{k!}$; using the property $\sum_{k=0}^{\infty} \frac{q^k}{k!} = e^q$; $\Rightarrow e^{-\lambda} e^{\lambda e^t}$
- Binomial MGF
  $E(e^{tk}) = \sum_{k=0}^{\infty} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{\infty} \binom{n}{k} (pe^t)^k (1-p)^{n-k} = [pe^t + (1-p)]^n$
- Exponential MGF
  $E(e^{tk}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{(t-\lambda)x} dx = \frac{\lambda}{\lambda - t}$, where $\lambda > t, \infty$ otherwise

4.4 Approximate Methods

For $Y = g(X)$

$$\mu_Y \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$$

$$\sigma_Y^2 \approx \sigma_X^2 [g'(\mu_X)]^2$$

5 Chapter 5: Limit Theorems

5.1 Definitions

5.1.1 Convergence in probability

$$\lim_{n\to\infty} P(|Z_n - \alpha| > \epsilon) = 0 \text{ for some } \alpha, \text{ any } \epsilon > 0$$

5.1.2 Almost sure convergence

$\forall\, \epsilon > 0,\ |Z_n - \alpha| > \epsilon$ only a finite number of times with $P = 1$
Summary: beyond some point in the sequence, the difference is always less than ϵ, but the location of that point is random.

5.2 Theorems

5.2.1 WLLN: weak law of large numbers

Let $\{X_i\}$ be sequence of iid RVs with $\mathrm{E}(X_i) = \mu,\ \mathrm{Var}(X_i) = \sigma^2$.
Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\forall\, \epsilon > 0$:

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \epsilon) = 0$$

Summary: $\overline{X}_n \xrightarrow{ip} \mu$

5.2.2 SLLN: strong law of large numbers

$\overline{X}_n \xrightarrow{as} \mu$

5.2.3 Continuity theorem

Let F_n be sequence of cdfs with mgfs M_n.
Let F be cdf with mgf M.

$$M_n(t) \to M(t)\ \forall t \text{ in an open interval containing } 0$$

$$\implies F_n \to F \text{ where } F \text{ cts}$$

5.2.4 CLT: central limit theorem

Let $\{X_i\}$ be sequence of iid RVs with $\mu = 0$, $\mathrm{Var} = \sigma^2$, common cdf F, mgf M defined about 0.
Let $S_n = \sum_{i=1}^{n} X_i$.

$$\implies \lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x)$$

$$\implies P\left(\frac{\overline{X}_n - \mathrm{E}(X)}{\sigma/\sqrt{n}} \le z\right) \to \Phi(z)$$

6 Chapter 6: Derivations from Normal

6.1 χ^2

6.1.1 χ^2_1

Let $Z \sim \mathcal{N}(0,1)$.

$$\implies U = Z^2 \sim \chi^2_1$$

$$\left(\frac{X-\mu}{\sigma}\right) \sim \mathcal{N}(0,1) \implies \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_1$$

Summary: square of normal RV is chi-squared, $df = 1$.

6.1.2 χ^2_n

Let $\{U_i\}_{i=1}^n$ iid χ^2_1.

$$\implies V = \sum_{i=1}^n U_i \sim \chi^2_n$$

Summary: sum of n chi-squared RVs is χ^2_n.

6.2 t

Definition:
Let $Z \sim \mathcal{N}(0,1)$, $U \sim \chi^2_n$, $Z \perp U$.

$$\implies \frac{Z}{\sqrt{U/n}} \sim t_n$$

Summary: t_n is normal RV divided by a scaled chi-squared with $df = n$

Density:

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

6.3 F

Definition:
Let U, V be iid χ^2 with $df = m, n$ respectively

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

Summary: F with $df = m, n$ found by dividing two chi-squared RVs divided by their dfs.

Density:

$$f(w) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(m/2)\,\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} w^{\frac{m}{2}-1}\left(1+\frac{m}{n}w\right)^{-\frac{(m+n)}{2}}$$

6.4 Sample Statistics

6.4.1 Definitions

Let $\{X_i\}_{i=1}^n$ be iid sample from $\mathcal{N}(\mu, \sigma^2)$.
<u>Sample mean</u>:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \mathrm{E}(\overline{X}) = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

<u>Sample variance</u>:

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

6.4.2 Theorems

- $\overline{X}, S^2$ independently distributed
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
- $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

7 Chapter 7: Sampling

- $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$
- $s^2_{\overline{X}} = \frac{s^2}{n}(1 - \frac{n}{N})$

8 Chapter 8: Estimation and Fitting

8.1 MoME: Method of Moments

8.1.1 Definitions

- $\underline{k^{th}\text{ moment}}$:

$$\mu_k = \mathrm{E}(X^k)$$

- $\underline{k^{th}\text{ sample moment}}$:
If $X_1, ..., X_n$ iid RVs, then

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

8.1.2 MoME

(1) Find low order moments; express moments in terms of parameters

(2) Find parameters in terms of moments

(3) Insert sample moments into expressions in (2)

8.2 MLE: Maximum Likelihood

8.2.1 Method of MLE

- $L(\theta) = f(\underline{x} \mid \theta)$
- $\ell(\theta) = \sum \ln[f(x_i \mid \theta)]$
- MLE maximizes ℓ

8.2.2 Large sample theory

- If f smooth, MLE from iid sample is consistent
- $I(\theta) = -\mathrm{E}(\ell'')$

8.2.3 MLE asymptotically unbiased

- Unbiased definition: $E(\hat{\theta}) = \theta$
- **Theorem**:
If f smooth, then $\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, 1)$
Summary: mle $\sim \mathcal{N}$ with $\mu = \theta_0$, asymptotic variance
- **Asymptotic variance**:

$$\mathrm{Var}(\theta_0) = \frac{1}{nI(\theta_0)} \approx -\frac{1}{\mathrm{E}(\ell'')}$$

8.2.4 CI for MLE

$$CI = \hat{\theta} \pm z_{\alpha/2} \cdot \sqrt{\mathrm{Var}(\theta_0)}$$

8.3 Bayes

8.3.1 Finding the posterior

$$f_{\Theta|X}(\theta \mid x) = \frac{f_{X,\Theta}(x, \theta)}{f_X(x)} = \frac{f_{X|\Theta}(x \mid \theta) f_\Theta(\theta)}{\int f_{X|\Theta}(x \mid \theta) f_\Theta(\theta)\, d\theta}$$

8.3.2 Bayesian paradigm

posterior $\propto$ likelihood $\cdot$ prior

$$f_{\Theta|X}(\theta \mid x) \propto f_{X|\Theta}(x \mid \theta) f_\Theta(\theta)$$

Conjugate Prior: The prior comes from one family G, the data conditional on G is of the family H, and the posterior is of the family G.

8.4 Consistent Estimate

Let $\hat{\theta}_n$ be an estimate of θ based on sample n.
Then, $\hat{\theta}_n$ <u>consistent</u> in probability if $\hat{\theta}_n \xrightarrow{ip} \theta$ as $n \to \infty$:

$$\forall\, \epsilon > 0,\ P(|\hat{\theta}_n - \theta| > \epsilon) \to 0 \ \text{ as } n \to \infty$$

8.5 Efficiency, CRLB

8.5.1 Efficiency

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

8.5.2 Cramer-Rao Inequality: CRLB

Let $\{X_i\}_{i=1}^n$ be iid with $f(x \mid \theta)$.
Let $T = t(X_1, ..., X_n)$ be unbiased estimator of θ. Then,

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

- If Var(T) = asymptotic variance, then <u>efficient</u>.
- MLE is <u>asymptotically efficient</u>.

8.6 Sufficiency

8.6.1 Sufficient statistic

$T(\underline{X})$ <u>sufficient</u> for θ if conditional distribution of $\underline{X}$ given $T = t$ does not depend on $\theta\ \forall\, t$
$\implies T$ is a <u>sufficient statistic</u>

8.6.2 Factorization theorem

T sufficient for $\theta \iff f(x_1, ..., x_n \mid \theta) = g[T(x_1, ..., x_n), \theta] \cdot h(x_1, ..., x_n)$

8.6.3 Exponential family

$$f(x \mid \theta) = e^{c(\theta)T(x) + d(\theta) + S(x)}$$

- T sufficient for $\theta \implies$ MLE $= f(T)$

8.6.4 Rao-Blackwell theorem

Let $\hat{\theta}$ be an estimator of θ with E$(\hat{\theta}^2)$ finite $\forall\, \theta$.
Suppose T is sufficient for θ, $\tilde{\theta} = \text{E}(\hat{\theta} \mid T)$.
Then, $\forall\, \theta$:

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

9 Chapter 9: Hypothesis Testing, Goodness of Fit

9.1 Definitions

- Null Hypothesis: H_0
- Alternate Hypothesis: H_A
- Type I Error: Rejecting H_0 when it is true
- Significance Level: α, probability of Type I Error
- Type II Error: Accepting H_0 when it is false
- Power: Probability of rejecting H_0 when it is false $= 1 - \beta$
- Simple Hypothesis: One vs. One
- Composite Hypothesis: One vs. Many
- P-value: smallest significance level at which the null hypothesis could be rejected

9.2 Likelihood Ratio (Test Statistic)

$$LR = \frac{P(x \mid X_0)}{P(x \mid H_1)} \cdot \frac{P(H_0)}{P(H_1)}$$

$\implies$ reject H_0 if $LR < c$

9.3 Neyman-Pearson Paradigm

9.3.1 Neyman-Pearson lemma

Suppose H_0, H_1 are *simple* hypotheses where test rejects H_0 when $LR < c$ with significance level α.
Then, any other test with significance level $\leq \alpha$ has power $\leq LR$ test.

9.3.2 UMP: Uniformly most powerful test

If H_1 composite, test that is most powerful $\forall$ simple alternatives in H_1 is uniformly most powerful (UMP)

9.4 Confidence Intervals

- **Confidence interval**:

$$P(\theta_0 \in C(X) \mid \theta = \theta_0) = 1 - \alpha$$

- **Acceptance region**

$$A(\theta_0) = \{X \mid \theta_0 \in C(X)\}$$

9.5 GLRT

9.5.1 Testing

$$\Lambda = \frac{\max_{\theta \in \omega_0} L(\theta)}{\max_{\theta \in \Omega} L(\theta)}$$

$\implies$ reject H_0 if $\Lambda < c$

9.5.2 GLRT Distribution Theorem

Under smoothness of pdfs, null distribution of $-2 \ln \Lambda \sim \chi^2_{df}$ with $df = \dim(\Omega) - \dim(\omega_0)$ as $n \to \infty$.

9.6 Multinomial Distribution

- **Hypothesis**: $\begin{cases} H_0 : p = p(\theta),\ \theta \in \omega_0 \\ H_1 : \text{ cell probabilities free} \end{cases}$

- **Distribution**:

$$\chi^2_{m-k-1} = \sum_{i=1}^{m} \frac{[x_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}$$

df = cells - num of estimated params - 1

9.7 Poisson Dispersion Test

- **Hypothesis**: $\begin{cases} H_0 : \text{Counts } x_1, ..., x_n \text{ Poisson with common } \lambda \\ H_1 : \text{Poisson with different rates} \end{cases}$

- **Result**:

$$-2\ln\Lambda = 2\sum_{i=1}^{n} x_i \ln\left(\frac{x_i}{\overline{x}}\right) \approx \frac{1}{\overline{x}}\sum_{i=1}^{n}(x_i - \overline{x})^2 \sim \chi^2_{n-1}$$

9.8 Hanging Rootograms

- Hanging histogram: n_j observed counts vs $\hat{n}_j$ predicted counts
  - variability not same across cells
- Hanging rootogram: $\sqrt{n_j} - \sqrt{\hat{n}_j}$
  - appx same variability
- Hanging chi-gram: $\frac{n_j - \hat{n}_j}{\sqrt{\hat{n}_j}}$
  - variance ≈ 1

9.9 Probability Plot

Plot of $F(X_{(k)})$ vs. $\frac{k}{n+1}$ OR plot of $X_{(k)}$ vs. $F^{-1}(\frac{k}{n+1})$

9.10 Tests for Normality

9.10.1 Coefficient of skewness

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^3}{s^3}$$

9.10.2 Coefficient of kurtosis

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4}$$

9.10.3 Variance-stabilizing transformation

$\text{Var}(Y) \approx \sigma^2(\mu)[f'(\mu)]^2$

10 Chapter 10: Summarizing Data

10.1 ecdf

10.1.1 Definition

Suppose $X_1, ...X_n$ sample/batch of iid numbers.

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_i)$$

10.1.2 Distribution

$nF_n(x) \sim \text{Binom}(n, F(x))$

- $\text{E}[F_n(x)] = F(x)$
- $\text{Var}[F_n(x)] = \frac{1}{n}F(x)[1 - F(x)]$

10.2 Survival Analysis

10.2.1 Survival function

$$S(t) = P(T > t) = 1 - F(t)$$

10.2.2 Hazard function

$$h(t) = \frac{f(t)}{1 - F(t)} = -\frac{d}{dt}\ln[1 - F(t)] = -\frac{d}{dt}\ln S(t)$$

10.3 QQ Plot

10.3.1 Definition

Plot quantiles of one distribution against vs. another where the quantiles are $x_p = F^{-1}(p)$

10.3.2 Common transformations

For control F and treatment G:

1. *Linear*: $y_p = x_p + h \implies G(y) = F(y - h)$
2. *Multiplicative*: $y_p = cx_p \implies G(y) = F(y/c)$

10.4 Kernel Density Estimate

Let w_h be a non-negative, symmetric weight function centered at 0 with $\int w = 1$.
Then, the kernel density estimate is:

$$f_h(x) = \frac{1}{n}\sum_{i=1}^{n} w_h(X - X_i)$$

- Represents a superposition of hills centered on the observations
- h = bandwidth: smoothness & bin width

10.5 Location

10.5.1 M estimates

- Sample mean minimizes negative log-likelihood, or the least squares estimate:

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

- Sample median minimizes:

$$\sum_{i=1}^{n} \left| \frac{X_i - \mu}{\sigma} \right|$$

- <u>M-estimate</u> minimizes:

$$\sum_{i=1}^{n} \Psi \left(\frac{X_i - \mu}{\sigma} \right)^2$$

11 Chapter 11: Comparing Two Samples

11.1 Two Independent Samples

11.1.1 Parametric: normal

1. **Overview**
   - Treatment: $X_1, ..., X_n$ iid $\mathcal{N}(\mu_X, \sigma^2)$
   - Control: $Y_1, ..., Y_n$ iid $\mathcal{N}(\mu_Y, \sigma^2)$
   - Pooled sample variance:

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2} = s^2_{\overline{X}-\overline{Y}}$$

   - **Thm**: distribution of difference

$$t = \frac{(\overline{X}-\overline{Y}) - (\mu_X - \mu_Y)}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{m+n-2}$$

2. **Hypothesis testing**
   - Hypothesis: $H_0 : \mu_X = \mu_Y$
   - Test statistic:

$$t = \frac{\overline{X}-\overline{Y}}{s_{\overline{X}-\overline{Y}}} \sim t_{m+n-2}$$

3. **Power**: power of rejecting H_0 when it is false
   - Factors that affect power:
     1) Real difference, $\Delta = |\mu_X - \mu_Y|$: large diff $\rightarrow$ greater power
     2) α: $\alpha \uparrow \implies$ power $\uparrow$
     3) σ: $\sigma \downarrow \implies$ power $\uparrow$
     4) Sample sizes n, m: $nm \uparrow \implies$ power $\uparrow$
   - Numerical power (assuming same n and variance from both samples):

$$1 - \Phi\left[z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right] + \Phi\left[-z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right]$$

11.1.2 Nonparametric: Mann-Whitney

1. **Overview**

   - H_0 : no treatment effect
   - U: sum of wins and ties in relevant set
   - T: total sum of ranks in set
   - Procedure:
     (1) Group all $m+n$ observations together, rank in order of increasing size
     (2) Calculate some of ranks of observations from control group
     (3) Reject H_0 if sum is too extreme

2. **Distribution version**

   - $X_1, ..., X_n \sim F$ control group
   - $Y_1, ..., Y_m \sim G$ experimental group
   - $H_0 : F = G$
   - **Thm**: for T_Y as rank sum of Y:

$$\mathrm{E}(T_Y) = \frac{m(m+n+1)}{2}$$

$$\mathrm{Var}(T_Y) = \frac{mn(m+n+1)}{12}$$

3. **Rank-sum version**

   - Mann-Whitney test statistic:

$$U_Y = T_Y - \frac{m(m+1)}{2}$$

   - **Thm**: under $H_0 : F = G$:

$$\mathrm{E}(U_Y) = \frac{mn}{2}$$

$$\mathrm{Var}(U_Y) = \frac{mn(m+n+1)}{12}$$

   - For m, n both > 10:

$$\frac{U_Y - \mathrm{E}(U_Y)}{\sqrt{\mathrm{Var}(U_Y)}} \sim \mathcal{N}(0,1)$$

11.1.3 Bayesian approach

1. **Assumptions**
   - X_i iid $\mathcal{N}$, mean μ_X, precision ξ
   - Y_j iid $\mathcal{N}$, mean μ_Y, precision ξ

2. **Procedure**
   (1) Assign prior to (μ_X, μ_Y, ξ)
   (2) Posterior $\propto$ prior $\times$ likelihood; normalize
   (3) Find marginal joint distribution by integrating out ξ
   (4) Find marginal for $\mu_X - \mu_Y$

3. **Approximate result**: use improper priors
   - Final posterior:

$$f_{post}(\mu_X, \mu_Y, \xi) \propto \xi^{\frac{n+m}{2}-1} \exp\left(-\frac{\xi}{2}\Big[(n-1)s_X^2 + (m-1)s_Y^2\Big]\right) \cdot \exp\left(-\frac{n\xi}{2}(\mu_X - \overline{x})^2\right) \cdot \exp\left(-\frac{m\xi}{2}(\mu_Y - \overline{y})^2\right)$$

   - Distributions:

$$\mu_X - \mu_Y \sim \mathcal{N}(\overline{X} - \overline{Y}, \sigma^2)$$

$$\sigma^2 = \xi^{-1}(n^{-1} + m^{-1})$$

   - Distribution of marginal posterior of $\mu_X - \mu_Y$:

$$\frac{\Delta - (\overline{X} - \overline{Y})}{s_{\overline{X}-\overline{Y}}} \sim t_{m+n-2}$$

4. **Bayes vs. frequentist**
   - Frequentist:
     - $\overline{X} - \overline{Y}, s_p$ random
     - $\Delta = \mu_X - \mu_Y$ fixed
   - Bayes:
     - $\overline{X} - \overline{Y}, s_p$ fixed
     - $\Delta = \mu_X - \mu_Y$ random
     - Statements about Δ from data

11.2 Paired Samples

11.2.1 Overview

1. **Assumptions**
   - Pairs (X_i, Y_i), $i = 1, ..., n$
   - Different pairs iid, but $\text{Cov}(X_i, Y_i) = \sigma_{XY}$
   - $D_i = X_i - Y_i$
2. **Population**
   - $\text{E}(D) = \mu_X - \mu_Y$
   - $\text{Var}(D) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$
3. **Estimates**
   - $\text{E}(\overline{D}) = \mu_X - \mu_Y$
   - $\text{Var}(\overline{D}) = \frac{1}{n}(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)$
4. **Simplification**: if $\sigma_X = \sigma_Y = \sigma$
   - $\text{Var}(\overline{D}) = \frac{2\sigma^2(1-\rho)}{n}$
   - $\text{Var}(\overline{D}_\perp) = \frac{2\sigma^2}{n}$
   - efficiency $= \frac{\text{Var}(\overline{D})}{\text{Var}(\overline{D}_\perp)} = 1 - \rho$

11.2.2 Parametric: normal/t-test

1. **Assumptions**
   - $X_i - Y_i$ sample from $\mathcal{N}$, $D = X - Y$
   - $\text{E}(D_i) = \mu_X - \mu_Y = \mu_D$
   - $\text{Var}(D_i) = \sigma_D^2$
2. **Inference**: σ_D unknown; $H_0 : \mu_D = 0$; ok for large n by CLT

$$t = \frac{\overline{D} - \mu_D}{s_{\overline{D}}} \sim t_{n-1}$$

11.2.3 Nonparametric: Signed-Rank Test

1. **Procedure**

   (1) Calculate differences D_i, find $|D_i|$, rank $|D_i|$

   (2) Restore signs of D_i to ranks to create signed ranks

   (3) Calculate $W_+ =$ sum of positive ranks as test statistic
2. **Test**
   - $H_0 : D_i$ distribution symmetric about 0
   - **Thm**: under H_0,

$$\text{E}(W_+) = \frac{n(n+1)}{4}$$

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

11.3 Experimental Design

- **Bonferroni method**: for multiple hypothesis testing, test each at α/n to achieve overall error of α

12 Chapter 12: ANOVA (F)

12.1 One-Way ANOVA

- **One-way layout**: independent measurements made under each of several treatments
- Sources of variability:
  1. Within samples
  2. Between samples

12.1.1 Normal theory: F-test

1. **Setup**
   - I = number of groups/treatments
   - J = sample size
   - $Y_{ij} = j^{th}$ observation of i^{th} treatment
2. **Model**: $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$
   - Variables:
     - μ = overall/total mean
     - α_i = differential effect of i^{th} treatment
     - ϵ_{ij} = random error in j^{th} observation of i^{th} treatment
   - Assumptions:
     - ϵ_{ij} iid $\mathcal{N}(0, \sigma^2)$
     - α_i normalized
3. **Sum of squares**
   - Notation:
     - $\overline{Y}_{i.} = \frac{1}{J}\sum_j Y_{ij}$
     - $\overline{Y}_{..} = \frac{1}{IJ}\sum_i \sum_j Y_{ij}$
   - Equation: $SS_{TOT} = SS_W + SS_B$
     - *Total sum of squares*: $SS_{TOT} = \sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2$
     - *Sum of squares within*: $SS_W = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2$
     - *Sum of squares between*: $SS_B = J\sum_i (Y_{i.} - \overline{Y}_{..})^2$
4. **Expected value theorems**
   - **Thm**: *expected SS*
     Let X_i be independent random variable with $\mathrm{E}(X_i) = \mu_i$, $\mathrm{Var}(X_i) = \sigma^2$. Then,
     $$\mathrm{E}(X_i - \overline{X})^2 = (\mu_i - \overline{\mu})^2 + \frac{n-1}{n}\sigma^2$$
     where $\overline{\mu} = \frac{1}{n}\sum_i \mu_i$
   - **Thm**: *expected value of SS_W, SS_B*
     $$\mathrm{E}(SS_W) = I(J-1)\sigma^2$$
     $$\mathrm{E}(SS_B) = J\sum_{i=1}^{I} \alpha_i^2 + (I-1)\sigma^2$$
5. **Variance rules & theorems**
   - Key observations:
     (1) SS_W can estimate σ^2: $s_p^2 = \frac{SS_W}{I(J-1)}$
     (2) If all $\alpha_i = 0$, $\frac{SS_W}{I(J-1)} \approx \frac{SS_B}{I-1}$; if some $\neq 0$, then SS_B inflated $\implies$ motivation for test

- **Thm**: *distribution of SS*
  If ϵ_{ij} iid $\mathcal{N}(0, \sigma^2)$:

  $$\frac{SS_W}{\sigma^2} \sim \chi^2_{I(J-1)}$$

  If also all $\alpha_i = 0$:

  $$\frac{SS_B}{\sigma^2} \sim \chi^2_{I-1}$$

  with $\frac{SS_W}{\sigma^2} \perp \frac{SS_B}{\sigma^2}$

6. **Test**

   - Test statistic: if H_0 true, $F \approx 1$

   $$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

   $$H_A : \alpha_i \neq 0$$

   $$F = \frac{SS_B/(I-1)}{SS_W/[I(J-1)]} \sim F_{I-1,\, I(J-1)}$$

7. **Test with different number of observations**: non-constant J_i

   (1) *The identity*

   $$\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 + \sum_i J_i(\overline{Y}_{i.} - \overline{Y}_{..})^2$$

   (2) *Expected values*

   $$\mathrm{E}(SS_W) = \sigma^2 \sum_i (J_i - 1)$$

   $$\mathrm{E}(SS_B) = \sum_{i=1}^{I} J_i \alpha_i^2 + (I-1)\sigma^2$$

8. **Summary**

   - The model: $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$
   - Assumptions:
     (1) $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$
       - F-test approximately valid for large enough samples even if non-normal
     (2) σ^2 CONSTANT
       - F-test not strongly affected by diff σ^2 as long as equal number of obs per group
     (3) ϵ_{ij} independent
       - Most important!!

9. **Tukey's method of multiple comparisons**

   - One-way anova: testing fact of difference, not measurement of difference or specific difference pairs
   - Tukey method: compare pairs/groups of treatment means via t-test
   - **Tukey test**: construct CIs for differences of all pairs of means such that intervals simultaneously have some set coverage probability; can use duality of CI/hypothesis testing to determine differences
   - *Assumptions*:
     - Sample sizes are equal (NOT required for Bonferroni)
     - $\epsilon \sim \mathcal{N}$ with constant σ^2

12.1.2 Nonparametric one-way: Kruskal-Wallis

1. **Setup**
   - *Assumptions*: independent observations, no necessary functional form
   - *Variables*:
     - R_{ij} = rank of Y_{ij} in pooled sample
     - $\overline{R}_{i.} = \frac{1}{J_i}\sum_{j=1}^{J_i} R_{ij}$: average rank in i^{th} group
     - $\overline{R}_{..} = \frac{N+1}{2}$
     - $SS_B = \sum_i J_i(\overline{R}_{i.} - \overline{R}_{..})^2$
2. **Test statistic**

$$K = \frac{12}{N(N+1)} SS_B = \frac{12}{N(N+1)}\left(\sum_{i=1}^{I} J_i \overline{R}_{i.}^2\right) - 3(N+1) \approx \chi^2_{I-1}$$

12.2 Two-Way ANOVA

- **Two-way anova**: experimental design involving two factors, each at 2+ levels
- *Assumptions*:
  - If I levels of f_1 and J levels of f_2, IJ combos
  - K independent observations taken from each combination (I, J)

12.2.1 Normal theory, 2-way

1. **Assumptions**
   - $K > 1$ observations per cell
   - <u>Balanced</u>: equal observations per cell
   - $Y_{ijk} = k^{th}$ observation in cell (i, j)
   - ϵ_{ijk} iid $\mathcal{N}(0, \sigma^2)$
2. **Model**
   - *The model*: $Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}$
   - *Constraints*:
     - *Row differential*: $\sum_i \alpha_i = 0$
     - *Column differential*: $\sum_j \beta_j = 0$
     - *Residual*: $\sum_i \delta_{ij} = \sum_j \delta_{ij} = 0$
3. **MLEs**
   - *Log-likelihood*:

$$\ell = -\frac{IJK}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})^2$$

   - *MLEs*:

$$\hat{\mu} = \overline{Y}_{...}$$

$$\hat{\alpha}_i = \overline{Y}_{i..} - \overline{Y}_{...}$$

$$\hat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$$

$$\hat{\delta}_{ij} = \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}$$

4. **SS**: $SS_{TOT} = SS_A + SS_B + SS_{AB} + SS_E$

$$SS_A = JK\sum_{i=1}^{I}(\overline{Y}_{i..}\overline{Y}_{...})^2$$

$$SS_B = IK\sum_{j=1}^{J}(\overline{Y}_{.j.}\overline{Y}_{...})^2$$

$$SS_{AB} = K\sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2$$

$$SS_E = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \overline{Y}_{ij.})^2$$

$$SS_{TOT} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \overline{Y}_{...})^2$$

5. **Expectations**

$$\mathrm{E}(SS_A) = (I-1)\sigma^2 + JK\sum_{i=1}^{I}\alpha_i^2$$

$$\mathrm{E}(SS_B) = (J-1)\sigma^2 + IK\sum_{j=1}^{J}\beta_j^2$$

$$\mathrm{E}(SS_{AB} = (I-1)(J-1)\sigma^2 + K\sum_{i=1}^{I}\sum_{j=1}^{J}\delta_{ij}^2$$

$$\mathrm{E}(SS_E) = IJ(K-1)\sigma^2$$

6. **Distributions of SS**

   (1) $\frac{SS_E}{\sigma^2} \sim \chi^2_{IJ(K-1)}$

   (2) Under $H_A : \alpha_i = 0$ for all i: $\frac{SS_A}{\sigma^2} \sim \chi^2_{I-1}$

   (3) Under $H_B : \beta_j = 0$ for all j: $\frac{SS_B}{\sigma^2} \sim \chi^2_{J-1}$

   (4) Under $H_{AB} : \delta_{ij} = 0$ for all i, j: $\frac{SS_{AB}}{\sigma^2} \sim \chi^2_{(I-1)(J-1)}$

   (5) SS are independently distributed

7. **The test**

   - Compare relevant SS to SS_E
   - F = ratio of MS where $MS = SS/df$; reject when $F \gg 1$
   - Example: *Interaction test*

$$F = \frac{SS_{AB}/[(I-1)(J-1)]}{SS_E/[IJ(K-1)]} = \frac{MS_{AB}}{MS_E}$$

12.2.2 Nonparametric: Friedman's test

- **Assumptions**: none on distribution: only according to ranks
- **Procedure**:

  (1) Within each of the J blocks, rank the observations

  (2) H_0 : no effect due to I treatments

  (3) Relevant variable: $SS_A = J\sum_{i=1}^{I}(\overline{R}_{i..} - \overline{R}_{...})^2$

  (4) Test statistic approximation:

$$Q = \frac{12J}{I(I+1)}SS_A \sim \chi^2_{I-1}$$

13 Chapter 13: Analysis of Categorical Data (χ^2)

- **Categorical data**: in counts from categories of two-way tables (contingency table)

13.1 Fisher's Exact Test

- *Test statistic*: N_{11}; hypergeometric under H_0
- *Probability*:

$$P(N_{11} = n_{11}) = \frac{\binom{n_{1.}}{n_{11}}\binom{n_{2.}}{n_{21}}}{\binom{n_{..}}{n_{.1}}}$$

13.2 Chi-Square Test of Homogeneity

1. **Setup**
   - Independent observations from J multinomial distributions, each of which has I cells/categories
   - *Test idea*: are all cell probabilities homoegeneous/equal (**goodness of fit test**)
   - π_{ij} = probability of i^{th} category in j^{th} multinomial
2. **Test**
   - $H_0 : \pi_{i1} = \pi_{i2} = \cdots = \pi_{iJ}$ for all i
   - n_{ij} = count in i^{th} category in j^{th} multinomial
3. **Thm**: *MLE of π's*
   - Under H_0, mle's of parameters π_i are:

     $$\hat{\pi}_i = \frac{n_{i.}}{n_{..}}$$

     - $n_{i.}$ = total responses in i^{th} category
     - $n_{..}$ = grand total responses
   - For j^{th} multinomial, expected count in i^{th} category:

     $$E_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}$$

     $$O_{ij} = n_{ij}$$
4. χ^2**-statistic**

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(I-1)(J-1)}$$

13.3 Chi-Square Test of Independence

1. **Setup**

   - Sample size n cross-classified in table with I rows, J columns contingency table
   - π_{ij} = joint distribution of n_{ij}
   - *Marginal probabilities*:

   $$\pi_{i.} = \sum_{j=1}^{J} \pi_{ij} \qquad \pi_{.j} = \sum_{i=1}^{I} \pi_{ij}$$

2. **Test**

   $$H_0 : \pi_{ij} = \pi_{i.}\pi_{.j}$$

3. **Thm**: *MLEs*

   $$H_0 : \hat{\pi}_{ij} = \hat{\pi}_{i.} + \hat{\pi}_{.j} = \left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.j}}{n}\right)$$

   $$H_1 : \hat{\pi}_{ij} = \frac{n_{ij}}{n}$$

13.4 Matched Pairs: McNemar's Test

1. **Test**: off-diagonal probabilities are equal

   $$H_0 : \pi_{12} = \pi_{21}$$

2. **MLEs**: under H_0:

   $$\hat{\pi_{11}} = \frac{n_{11}}{n}$$

   $$\hat{\pi_{22}} = \frac{n_{22}}{n}$$

   $$\hat{\pi_{12}} = \hat{\pi_{21}} = \frac{n_{12} + n_{21}}{n}$$

3. **Test statistic**

   $$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \sim \chi_1^2$$

13.5 Odds Ratio

1. **Definitions**

   - **Odds**:

   $$odds(A) = \frac{P(A)}{1 - P(A)}$$

   - **Odds ratio**: influence of X on D:

   $$\Delta = \frac{odds(D \mid X)}{odds(D \mid X^C)} = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}} = \frac{\text{product of diag probs}}{\text{product of off-diag probs}}$$

2. **Sampling methods**

   (1) Random sample from entire population:
   - If D rare, need large n to guarantee enough D

   (2) Prospective study: fixed number of X, X^C sampled; compare incidence of D in the groups
   - Can compare & estimate $P(D \mid X)$, $P(D \mid X^C)$ and odds ratio
   - Individual probabilities π_{ij} cannot be estimated because marginal counts fixed

   (3) Retrospective study: fixed number of D, D^C sampled; compare incidence of X in the groups
   - Can directly estimate $P(X \mid D)$, $P(X \mid D^C)$
   - Can't estimate $P(D \mid X)$, $P(D \mid X^C)$ since marginal counts fixed
   - Same odds ratio Δ
   - Estimate: $\hat{\Delta} = \frac{n_{00}n_{11}}{n_{10}n_{01}}$