Definition

Bonferroni's Inequality

Fundamentals of Probability

Definition

Bayes' Rule

Fundamentals of Probability

Definition

Sensitivity and Specificity

Fundamentals of Probability

$$P(A,B) \geq P(A) + P(B) - 1$$

This is useful if you are asked to give the minimum of P(A,B)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Sensitivity=$P(T=1|D=1)$

Specificity=$P(T=0|D=0)$

Equation

Change of Variable

Fundamentals of Probability

Equation

Moment Generating Function

Fundamentals of Probability

Equation

Location-Scale shift

Fundamentals of Probability

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(g)\right|$$

Note that this only works for 1-1 monotonic functions

$$E[e^{xt}] = \int_x e^{xt} f(x)dx$$

Then evaluate the derivative for each moment at $t = 0$. For example, the second moment would be the second derivative of the mgf evaluated at $t = 0$.

$$f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$$

Basically, multiply the pdf by $\frac{1}{\sigma}$ and replace z with $\frac{x-\mu}{\sigma}$

Form

Exponential Family

Pro Tip

Mgf when $X \perp Y$

Equation

Bivariate Transformations

$$f(x|\theta) = h(x)c(\theta)\exp\left\{\sum_{i=1}^{k} w_i(\theta)t_i(x)\right\}$$

If a family is exponential and there is a non-empty parameter space, it is considered "complete". Therefore $t(x)$ is MSS.

$$Mgf(x+y) = Mgf(x)Mgf(y)$$

This is cool because it shows that if $X \perp Y$ and $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, then $X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

$$f_{u,v}(u,v) = f_{x,y}(h_1(u,v), h_2(u,v))|J|$$

This is for continuous. DON'T FORGET THE JACOBIAN. or Jacob

Marley will come after you.

Equation

Iterative Expectations

Fundamentals of Probability

Equation

Iterative Variance

Fundamentals of Probability

Definition

Covariance

Fundamentals of Probability

$$E[Y] = E_X[E_{Y|X}[Y|X]]$$

$$Var[Y] = E_X[Var[Y|X]] + Var_X[E[Y|X]]$$

$$\begin{aligned} cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E[XY] - \mu_X\mu_Y \end{aligned}$$

Definition

Correlation

Definition

Chebyshev's Inequality

Definition

Jensen's Inequality

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

$$P(g(X) \geq r) \leq \frac{E(g(X))}{r}$$

$$P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$$

$$E(g(x)) \geq g(E(x))$$

This holds if $g(x)$ is a convex function. That means it has a positive 2nd derivative, like a smile. I have high expectations!

Definition

Holder's Inequality

Fundamentals of Probability

Proof

Correlation is Bounded by -1 and 1

Fundamentals of Probability

Equation

Order Statistics PDF

Fundamentals of Probability

$$|E(XY)| \leq E|XY| \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}$$
$$\frac{1}{p} + \frac{1}{q} = 1$$

Cauchy-Schwartz is a special case where $p = q = 2$. Can use to show correlation is bounded by -1 and 1.

$$\begin{aligned}
\Big|cov(X,Y)\Big| &= \Big|E[(X-\mu_x)(Y-\mu_y)]\Big| \\
&\leq (E|X-\mu_x|^2)^{1/2}(E|Y-\mu_y|^2)^{1/2} \leq \sqrt{\sigma_x^2}\sqrt{\sigma_y^2} \\
&\leq \sigma_x\sigma_y
\end{aligned}$$
$$\Big|corr(X,Y)\Big| = \Big|\rho\Big| = \Big|\frac{cov(X,Y)}{\sigma_x\sigma_y}\Big| = \frac{\Big|cov(X,Y)\Big|}{\sigma_x\sigma_y} \leq \frac{\sigma_x\sigma_y}{\sigma_x\sigma_y} = 1$$

$$\frac{n!}{(j-1)!(n-j)!} f(x)[F(x)]^{j-1}[1-F(x)]^{n-j}$$

Equation

Order Statistics CDF

Equation

Joint PDF of Order Statistics

Pro Tip

If $X_i \overset{iid}{\sim} Unif(0,1)$ *the pdf of the kth order statistic*

$$\sum_{i=j}^{n} \binom{n}{i} F(x)^i [1 - F(x)]^{n-i}$$

$$\frac{n!}{(l-1)!(m-l-1)!(n-m)!}$$
$$F(x_l)^{l-1} f(x_l) f(x_m) [F(x_m) - F(x_l)]^{m-l-1} [1 - F(x_m)]^{n-m}$$

soooo its so long it can't fit on one line...but this is all one equation. memorize it fooooool.

$$Beta(k, n-k+1)$$

Proof

Convergence in Probability

Definition

Convergence Almost Surely

Definition

Convergence in Distribution

Convergence in probability means that the estimator is consistent. We can prove something converges in probability using Chebychev's. Also, the Weak Law of Large Numbers is that $\bar{X} \xrightarrow{p} \mu$.

$$\begin{aligned} P(|\bar{X}_n - X| \geq \epsilon) &\leq \frac{E[(\bar{X}_n - \mu)^2]}{\epsilon^2} \\ &\leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \\ \lim_{n\to\infty} P(|\bar{X}_n - X| \geq \epsilon) &\leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} \\ \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} &= 0 \\ \lim_{n\to\infty} P(|\bar{X}_n - X| \geq \epsilon) &= 0 \end{aligned}$$

$$P(\lim_{n\to\infty} |\bar{X}_n - X| \geq \epsilon) = 0$$

Convergence almost surely implies convergence in probability which implies convergence in distribution.

$$X_n \xrightarrow{d} X \iff \lim_{n\to\infty} F_{x_n}(x) = F_x(x) \text{ for all x where } F_x(x) \text{ is continuous}$$

The CLT is convergence in distribution. Coming to a flashcard near you - proof of the CLT using MGFs. Yeah. You can't wait.

Definition

Slutsky's

Definition

Delta Method

Definition

Accept-Reject Algorithm

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$ then:
$Y_n X_n \xrightarrow{d} aX$
and $Y_n + X_n \xrightarrow{d} a + X$

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \frac{\sigma}{S_n} \xrightarrow{d} N(0, 1)$$

because

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$
$$\frac{\sigma}{S_n} \xrightarrow{p} 1$$

1st order:

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2 var(X_n))$$

2nd order:

$$n(g(X_n) - g(\theta)) \xrightarrow{d} \frac{Var(X_n)}{2} g''(\theta)\chi_1^2$$

To generate Y from $f(y)$:

1. Generate v from a known distribution, $f(v)$, with support that contains the support of Y.
2. Calculate $M = \sup_y \frac{f_Y(y)}{f_V(y)}$ (the supremum of Y in V)
3. Generate $U \sim Unif(0, 1)$ if $U \leq \frac{1}{M}\frac{f_Y(v)}{f_V(v)}$ then accept, otherwise reject.

Distribution

Bernoulli

Fundamentals of Probability

Distribution

Binomial

Fundamentals of Probability

Distribution

Geometric

Fundamentals of Probability

$$
\begin{aligned}
X &\sim Bin(1,p)\\
f(x) &= p^x(1-p)^{1-x} \text{ for } x = 0,1\\
E(x) &= p\\
Var(x) &= p(1-p)
\end{aligned}
$$

$$
\begin{aligned}
X &\sim Bin(n,p)\\
f(x) &= \binom{n}{x} p^x(1-p)^{n-x}\\
E(x) &= np\\
Var(x) &= np(1-p)
\end{aligned}
$$

$$
\begin{aligned}
X &\sim Geometric(p) \text{ or } X \sim NegBinom(1,p)\\
f(x) &= p(1-p)^{x-1} \text{ for x=1,...}\\
E(x) &= \frac{1}{p}\\
Var(x) &= \frac{1-p}{p^2}
\end{aligned}
$$

Distribution

Negative Binomial

Fundamentals of Probability

Distribution

Hypergeometric

Fundamentals of Probability

Distribution

Discrete Uniform

Fundamentals of Probability

$$
\begin{aligned}
X &\sim NegBinom(r,p) \\
f(x) &= \binom{x-1}{r-1} p^r (1-p)^{x-r} \text{ for x=r,r+1,...} \\
E(x) &= \frac{r}{p} \\
Var(x) &= \frac{r(1-p)}{p^2}
\end{aligned}
$$

where x is the number of experiments needed to get r successes

$N = \#$ of balls, $K = \#$ selected $M = \#$ of successes $X = \#$ of successes in your sample

$$
\begin{aligned}
f(x) &= \frac{\binom{M}{X}\binom{N-M}{K-X}}{\binom{N}{K}} \\
E(x) &= \frac{KM}{N} \\
Var(x) &= \frac{KM}{N}\left(\frac{N-M}{N}\right)\left(\frac{N-K}{N-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
X &\sim DUnif(a,b) \\
f(x) &= \frac{1}{b-a+1} \text{ for x=a,...,b} \\
E(x) &= \frac{a+b}{2} \\
Var(x) &= \frac{(b-a+1)^2-1}{12}
\end{aligned}
$$

Distribution

Poisson

Fundamentals of Probability

Distribution

Uniform

Fundamentals of Probability

Distribution

Gamma

Fundamentals of Probability

$$X \sim Pois(\lambda)$$
$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for x=0,1,2...}$$
$$E(x) = \lambda$$
$$Var(x) = \lambda$$

$$X \sim Unif(a, b)$$
$$f(x) = \frac{1}{b-a} \text{ a<x<b}$$
$$E(x) = \frac{a+b}{2}$$
$$Var(x) = \frac{(b-a)^2}{12}$$

$$X \sim Gamma(\alpha, \beta)$$
$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \text{ for } x > 0,\ \alpha > 0,\ \beta > 0$$
$$E(x) = \alpha\beta$$
$$Var(x) = \alpha\beta^2$$

Distribution

Chi-square

Fundamentals of Probability

Distribution

Exponential

Fundamentals of Probability

Distribution

Normal

Fundamentals of Probability

$$
\begin{aligned}
X &\sim \chi^2(p) \\
f(x) &= \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2} \\
E(x) &= p \\
Var(x) &= 2p
\end{aligned}
$$

special case of Gamma where $\alpha = p/2$ and $\beta = 2$.

$$
\begin{aligned}
X &\sim Exp(\beta) \\
f(x) &= \beta e^{-x\beta} \text{ for } x > 0,\ \beta > 0 \\
E(x) &= \frac{1}{\beta} \\
Var(x) &= \frac{1}{\beta^2}
\end{aligned}
$$

Special case of Gamma where $\alpha = 1$

$$
\begin{aligned}
X &\sim N(\mu, \sigma^2) \\
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \text{ for } -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0 \\
E(x) &= \mu \\
Var(x) &= \sigma^2
\end{aligned}
$$

Distribution

Beta

Distribution

Cauchy

Pro Tip

e^t

$$X \sim Beta(\alpha, \beta)$$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } 0 < x < 1,\ \alpha > 0,\ \beta > 0$$

$$E(x) = \frac{\alpha}{\alpha+\beta}$$

$$Var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$X \sim Cauchy(\theta)$$

$$f(x) = \frac{1}{\pi(1+(x-\theta)^2)} \text{ for } -\infty < x < \infty,\ -\infty < \theta < \infty$$

$$\lim_{n\to\infty}\left(1+\frac{t}{n}\right)^n$$

Pro Tip

Binomial Series

Fundamentals of Probability

Pro Tip

Geometric Series

Fundamentals of Probability

Definition

Positive Predictive Value

Statistical Inference

$$\sum_{v=0}^{u} \frac{u!}{(u-v)!v!} \theta^{u-v} \lambda^{v} = (\theta + \lambda)^{u}$$

$$\sum_{i=1}^{n} q^{i-1} = \frac{1-q^{n}}{1-q}$$

$$P(D^{+}|T^{+}) = \left[1 + \frac{P(T^{+}|D^{-})}{P(T^{+}|D^{+})} \frac{P(D^{-})}{P(D^{+})}\right]^{-1}$$

Definition

Kullback-Leibler Divergence

Definition

Hellinger Distance

Proof

LR is bounded by $1/k$

$$KLD(g,f) = E_g\left[\log\frac{g(X)}{f(X)}\right] \geq 0$$

Measures how much information you lose by using the worse distribution.

$$KLD(g,f) \geq 2[H(f,g)]^2$$

Lower bound for the KLD

$$P_g\left(\frac{\prod f(x_i)}{\prod g(x_i)} > k\right) \leq \frac{E_g\left[\frac{\prod f(x_i)}{\prod g(x_i)}\right]}{k} = \frac{1}{k}$$

$$\begin{aligned} E_g\left[\frac{\prod f(x_i)}{\prod g(x_i)}\right] &= \int \frac{\prod f(x_i)}{\prod g(x_i)} \prod g(x_i)dx \\ &= \int \prod f(x_i)dx = \int P_f(X_1, ..., X_n)dx = 1 \end{aligned}$$

Proof by Markov's inequality. This bound holds if we look at the data as it accumulates.

PROOF

Asymptotic behavior of LR

PROOF

Convergence of the posterior

DEFINITION

Rational for Maximum Likelihood

As evidence accumulates, the LR converges to 0

$$LR_n = \exp\left\{\log \prod \frac{f(x_i)}{g(x_i)}\right\} = \exp\left\{\sum \log f(x_i) - \sum \log g(x_i)\right\}$$
$$= \exp\left\{n\left[\frac{1}{n}\sum \log f(x_i) - \frac{1}{n}\sum \log g(x_i)\right]\right\}$$

From the LLN: $\frac{1}{n}\sum \log \frac{f(x_i)}{g(x_i)} \to E_g\left[\log \frac{f(x_i)}{g(x_i)}\right] \leq \log E_g\left[\frac{f(x_i)}{g(x_i)}\right]$ by Jensen's

$$\log E_g\left[\frac{f(x_i)}{g(x_i)}\right] = \log(1) = 0$$

Therefore this portion some negative number, call it -c

$$\prod \frac{f(x_i)}{g(x_i)} \to \lim_{n\to\infty} e^{n[-c]} = 0$$

Due to the LR convergence properties, the posterior converges as well. (Note this proof is in the discrete case).

$$X_1, ..., X_n \overset{iid}{\sim} f(X;\theta_0)$$
$$P(\theta = \theta_0|\underline{\mathrm{X}}) = \left[1 + \sum_{\theta\neq\theta_0} \frac{P(\underline{\mathrm{X}}|\theta)}{P(\underline{\mathrm{X}}|\theta_0)}\frac{P(\theta)}{P(\theta_0)}\right]^{-1}$$
$$\frac{P(\underline{\mathrm{X}}|\theta)}{P(\underline{\mathrm{X}}|\theta_0)} \to 0 \text{ By LR convergence principle}$$
$$\frac{P(\underline{\mathrm{X}}|\theta)}{P(\underline{\mathrm{X}}|\theta_0)}\frac{P(\theta)}{P(\theta_0)} \to 0 \text{ as } n\to\infty \text{ therefore } \sum_{\theta\neq\theta_0}\frac{P(\underline{\mathrm{X}}|\theta)}{P(\underline{\mathrm{X}}|\theta_0)}\frac{P(\theta)}{P(\theta_0)} \to 0$$
$$P(H_0|\underline{\mathrm{X}}) \to 1 \text{ as } n\to\infty$$

Because $\hat{\theta}$ maximizes the likelihood function, it is the parameter value that is best supported by the data by the Law of Likelihood.

Definition

Invariance of the MLE

Definition

Bias

Definition

Variance

If $\hat{\theta}$ is the MLE for θ then $g(\hat{\theta})$ is the MLE for $g(\theta)$ as long as $g(\theta)$ is a 1-1 function of θ.

$$E[\hat{\theta} - \theta] = b(\hat{\theta})$$

$$E[(\hat{\theta} - E[\hat{\theta}])^2]$$

Definition

MSE

Definition

Consistency

Definition

Biases of MLEs

$$E[(\hat{\theta} - \theta)^2]$$
$$= Var[\hat{\theta}] + b^2(\hat{\theta})$$

$$\hat{\theta} \to \theta \text{ as } n \to \infty \text{ in probability, a.s., etc.}$$

This implies that the limiting bias is 0.

MLEs are often biased. For example:

- The MLE of the variance in the Normal case has a slight **negative** bias $-\frac{\sigma^2}{n}$. This goes to 0 in large samples
- Poisson mean inverse - the bias is undefined! zabert alert!

Definition

Bayes Estimator

Statistical Inference

Proof

Consistency of MLEs

Statistical Inference

Pro Tip

When will the MLE not be consistent?

Statistical Inference

Trade some bias for a reduction in variance.

$$f(\theta|\underline{\mathrm{x}}) = \frac{f(\underline{\mathrm{X}}|\theta)f(\theta)}{\int_{\Theta} f(\underline{\mathrm{X}}|\theta)f(\theta)d\theta}$$

Here, a posterior mean is achieved by shrinking the sample mean towards the prior mean.

To show consistency:

$$\hat{\theta}_n \xrightarrow{p} \theta \text{ as } n \to \infty$$

$$\text{In other words } \hat{\theta}_n - \theta = o_p(1)$$

$$\textbf{Method 1 } P(|\hat{\theta}_n - \theta) \geq \epsilon) \to 0$$

$$\textbf{Method 2 } \text{quadratic mean } MSE(\hat{\theta}_n) = Var(\hat{\theta}_n) + b^2(\hat{\theta}_n) \to 0$$

If you can show that bias$\to$ 0 and var$\to$ 0 then $\hat{\theta}_n \xrightarrow{qm} \theta$ which implies $\hat{\theta}_n \xrightarrow{p} \theta$.

When the number of parameters is increasing as $n \to \infty$. Here is an example where the MLE is not consistent from Neyman-Scott:

$$\begin{aligned}
Y_{11}, Y_{12} &\sim N(\mu_1, \sigma^2) \\
Y_{21}, Y_{22} &\sim N(\mu_2, \sigma^2) \\
. &\sim . \\
. &\sim . \\
Y_{n1}, Y_{n2} &\sim N(\mu_n, \sigma^2) \\
\hat{\sigma^2} &= \sum_{i=1}^{n} \sum j = 1^2 \frac{(Y_{ij} - \bar{Y}_i)^2}{2n}
\end{aligned}$$

$$\hat{\sigma}^2 \xrightarrow{p} \frac{\sigma^2}{2}$$

Definition

Continuous Mapping Theorem

Definition

Conditions for MLE consistency

Definition

Score Function

This is basically the best theorem out there.

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$
$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$
$$X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

Obviously the function has to be continuous for this to work.

1. Identifiability
2. Compactness of the parameter space (it is sufficient to assume concavity of the log LF and MLE cannot be at the boundary of the parameter space)
3. Continuity of $L(\theta)$ in θ - to ensure smoothness and existence of derivatives
4. Dominance: $|\log f(x;\theta)| < D(x)\ \forall \theta \in \Theta$

- First derivative of the log-likelihood function
- Unbiased estimator of zero

Definition

Fisher's Information

Definition

Bartlett's Second Identity

Proof

Asymptotic Normality of MLE

Information is the variance of the score function.

$$\mathcal{I}(\theta) = Var(S_i) = E[S_i^2]$$
$$\mathcal{I}_n(\theta) = Var\left(\sum S_i\right) = n\mathcal{I}(\theta)$$

It can be estimated by:

$$\frac{\sum S_i^2}{n} = \frac{1}{n}\sum_i \left(\frac{\partial \log f(x_i;\theta)}{\partial \theta}\right)^2$$

Under the correct model:

$$Var(S_i) = E[S_i^2] = -E[S_i']$$

$$l_i = logf(x_i;\theta)$$

By Taylor Series Expansion:

$$0 = l_n'(\hat{\theta}_n) \approx l_n'(\theta) + (\hat{\theta}_n - theta)l_n''(\theta) + R_n$$
$$(\hat{\theta}_n - \theta) \approx \frac{l_n'(\theta)}{-l_n''(\theta)} \Rightarrow \sqrt{n}(\hat{\theta}_n - \theta) \approx \frac{\frac{1}{\sqrt{n}}l_n'(\theta)}{-\frac{1}{n}l_n''(\theta)}$$
$$\sqrt{n}\frac{1}{n}\sum l_i'(\theta) \xrightarrow{d} N(0,\mathcal{I}(\theta)) \text{ by CLT and } -\frac{1}{n}l_n''(\theta) \xrightarrow{p} \mathcal{I}(\theta)\text{by LLN}$$

By Slutsky's $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \frac{1}{\mathcal{I}(\theta)})$

$$\sqrt{n\mathcal{I}(\hat{\theta}_n)}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1)$$

Definition

What happens to the MLE when the working model fails?

Definition

Asymptotic Normality of the MLE under Model Failure

Equation

Making a likelihood robust

The MLE $\hat{\theta}_n$ converges to θ_g where:

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{\sum_i \log f(x_i;\theta)}{n} \to \underset{\theta \in \Theta}{\operatorname{argmax}} E_g[\log f(x_i;\theta)] = \theta_g$$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, a^{-1}ba^{-1}) \text{ as } n \to \infty$$
$$\sqrt{\frac{\hat{a}^2 n}{\hat{b}}}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1)$$

You can make a likelihood robust by:

$$L_R(\theta) = L(\theta)^{\hat{a}/\hat{b}}$$

$$L_R(\theta) = L(\theta)^{\hat{a}/\hat{b}}$$
$$\hat{a} = -\frac{1}{n}\sum \frac{\partial^2 \log f(x_i;\hat{\theta}_n)}{\partial \theta^2}$$
$$\hat{b} = \frac{1}{n}\sum \left(\frac{\partial \log f(x_i;\hat{\theta}_n)}{\partial \theta}\right)^2$$

Definition

Unbiased Estimating Equation

Definition

Standardized

Examples

Natural Estimating Equations

$$E[g(\underline{\mathrm{X}};\theta)] = 0 \ \forall\theta \in \Theta$$

$$g_s(\underline{\mathrm{X}};\theta) = \frac{g(\underline{\mathrm{X}};\theta)}{E\left[\frac{\partial g(\underline{\mathrm{X}};\theta)}{\partial\theta}\right]} \ \forall\theta \in \Theta$$

- score functions
- equations from MOM estimation

Theorem

Optimality of the Score Function

Definition

Variance of the Standardized Score Function

Definition

What form does the estimating equation have to be to achieve the variance lower bound?

This is literally the Godambe Theorem of 1960. NOT A JOKE (but what an awesome name!)

1. The variance of a standardized estimating equation is bounded below by $1/\mathcal{I}_n(\theta)$,

$$Var[g_s(\underline{X};\theta)] = \frac{E_\theta[g^2]}{\left\{E_\theta\left[\frac{\partial g}{\partial \theta}\right]\right\}^2} \geq \frac{1}{E_\theta\left[\left(\frac{\partial \log f}{\partial \theta}\right)^2\right]}$$

2. It follows that $\forall\ g \in G$,

$$\frac{E_\theta[g^2]}{\left\{E_\theta\left[\frac{\partial g}{\partial \theta}\right]\right\}^2} \geq \frac{E_\theta[(g^*)^2]}{\left\{E_\theta\left[\frac{\partial g^*}{\partial \theta}\right]\right\}^2}$$

$$\frac{1}{\mathcal{I}_n(\theta)}$$

$$g(\underline{X};\theta) = a(\theta)\left\{T(\underline{X}) - \underbrace{E_\theta[T(\underline{X})]}_{h(\theta)}\right\}$$

This implies that $T(\underline{X})$ is the best unbiased estimator for $h(\theta)$. This achieves the CRLB.

Definition

Cramer-Rao Lower Bound

Statistical Inference

Definition

Sufficient Statistic

Statistical Inference

Pro Tip

The most famous MSS (oh yeahhh!)

Statistical Inference

$$Var[T(\underline{X})] \geq \frac{\{h'(\theta)\}^2}{\mathcal{I}_n(\theta)}$$

This is the smallest possible variance for any unbiased estimator of $h(\theta)$

$$f_{\underline{X}}(\underline{X};\theta) = g(T(\underline{X});\theta)h(\underline{X})$$

If the pdf can be factorized as above, then $T(\underline{X})$ is a sufficient statistic for θ

The likelihood function. le duh, whose class is this any ways?

Definition

Minimal Sufficient Statistic

Pro Tip

Technique to find the MSS

Definition

Rao-Blackwellization

A sufficient statistic is minimally sufficient if it is a function of every other sufficient statistic.

$$\frac{f(\underline{x};\theta)}{f(\underline{y};\theta)} = c(\underline{x}, \underline{y})$$

Here $(\underline{x}, \underline{y})$ is free of θ. So you find a way to set these equal to cross all of the θs out.

Conditioning on a sufficient statistic always yields a better estimator, with a variance less than or equal to that of the first estimator.

Definition

Ancillary Statistic

Statistical Inference

Definition

Completeness

Statistical Inference

Definition

Basu's Theorem

Statistical Inference

A statistic is ancillary if its distribution does not depend on θ.

A family is complete if

$$E_\theta(g(t)) = 0 \ \forall\theta \Rightarrow P_\theta(g(t) = 0) = 1 \ \forall\theta$$

Exponential families with non-empty parameter space are complete.

If $T(\underline{\mathrm{X}})$ is complete and a minimally sufficient statistic, then $T(\underline{\mathrm{X}})$ is independent of every ancillary statistic.

DEFINITION

MSS/CSS Lemma

DEFINITION

Lehmann-Scheffe Theorem

DEFINITION

Checking for completeness

If a MSS exists, than any CSS is also the MSS.

If $T(\underline{X})$ is a CSS (and therefore a MSS), then any statistic $h[T(\underline{X})]$ with fine variance is the MVUE of its expectation $E[h[T(\underline{X})]]$. In other words if an estimator is a function of a CSS, then it has the smallest variance among all estimators of its expected value.

1. Exponential families are complete as longs as the interior of the parameter space is non-empty
2. A sufficient statistic $T(\underline{X})$ are complete if no function is first order ancillary.

Proof

Lehmann-Scheffe Theorem

Statistical Inference

Proof

Uniqueness of the MVUE

Statistical Inference

Pro Tip

Conditionality Principle

Statistical Inference

Proof by contradiction. Suppose $h[T(\underline{X})]$ is unbiased for γ and $h[T(\underline{X})]$ is not the MVUE of γ. Then there exists another estimator, say $W(\underline{X})$ such that $E[W(\underline{X})] = \gamma$ and $Var[W(\underline{X})] < Var[h[T(\underline{X})]]$
Using Rao-Blackwellization, we can create a new estimator $r[T(\underline{X})] = E[W(\underline{X})|T(\underline{X})]$ such that $E[r[T(\underline{X})]] = \gamma$ and

$$Var[r[T(\underline{X})]] < Var[W(\underline{X})] < Var[h[T(\underline{X})]]$$

Notice both $r[T(\underline{X})]$ and $h[T(\underline{X})]$ are unbiased for γ, so $E[r[T(\underline{X})] - h[T(\underline{X})]] = 0 \ \forall \gamma$
But completeness implies that $r[T(\underline{X})] = h[T(\underline{X})]$ with probability 1, so we must have

$$Var[r[T(\underline{X})]] = Var[h[T(\underline{X})]]$$

Which contradicts the previous inequality. This completes the proof that an unbiased function of the CSS is the MVUE.

If $T(\underline{X})$ and $S(\underline{X})$ are MVUE for γ then $E[T(\underline{X})] = E[S(\underline{X})] = E\left[\frac{T(\underline{X})+S(\underline{X})}{2}\right]$. It follows that $Var[T(\underline{X})] = Var[S(\underline{X})]$ but

$$\begin{aligned} Var\left[\frac{T(\underline{X})+S(\underline{X})}{2}\right] &= \frac{1}{4}\left[Var[T(\underline{X})] + Var[S(\underline{X})] + 2Cov[T(\underline{X}), S(\underline{X})]\right] \\ &= \frac{1}{4}[2Var[S(\underline{X})] + 2\rho Var[S(\underline{X})]] = Var[S(\underline{X})]\left(\frac{1+\rho}{2}\right) \end{aligned}$$

This implies that $Var\left[\frac{T(\underline{X})+S(\underline{X})}{2}\right] \leq Var[S(\underline{X})]$. Because $S(\underline{X})$ is the MVUE, this must be an equality and $Var\left[\frac{T(\underline{X})+S(\underline{X})}{2}\right] = Var[S(\underline{X})]$. By **Cauchy-Schwartz inequality**, this equality only holds when $S(\underline{X}) = aT(\underline{X}) + b$. We know that $E[T(\underline{X})] = E[S(\underline{X})]$ so $a = 1$ and $b = 0$, therefore P(T($\underline{X}$)=S($\underline{X}$))=1

Always condition on Ancillary Statistics!

Pro Tip

Likelihood Prinicple

Statistical Inference

Definition

Criteria for Confidence Intervals

Statistical Inference

Proof

Quantile Convergence

Statistical Inference

If two experiments yield likelihood functions that are proportional, then those two sets of data are equivalent as statistical evidence. If likelihoods are the same, evidence should be the same. Inferences can of course be different.

1. Consistent estimator of the parameter: $\hat{\theta}_n \xrightarrow{p} \theta$
2. Asymptotic Normality: $\sqrt{\mathcal{I}_n(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1)$
3. Consistent estimator of the information: $\frac{\mathcal{I}_n(\hat{\theta}_n)}{\mathcal{I}_n(\theta)} \xrightarrow{p} 1$
4. (also) Expected Length
5. Unbiasedness
6. Selectivity

Proof by contradiction. Assume that $C_n \not\to Z$ the either

1. $\exists \delta$ s.t. $\forall n \; \exists \; n' > n \Rightarrow C_{n'} > Z + \delta$
2. $\exists \delta$ s.t. $\forall n \; \exists \; n' > n \Rightarrow C_{n'} < Z - \delta$

If 1, then $\forall n \; \exists n' > n$ s.t. $F_{n'}(C_{n'}) \geq F_{n'}(Z + \delta)$ and
$\therefore \lim_{n' \to \infty} F_{n'}(C_{n'}) \geq \lim_{n' \to \infty} F_{n'}(Z + \delta) = F(Z + \delta) > F(Z)$

If 2, then $\therefore \lim_{n' \to \infty} F_{n'}(C_{n'}) \leq F(Z - \delta) < F(Z)$

However, we know that $F_n(C_n) = \alpha \; \forall n$ (by definition) and $F(Z) = \alpha$. So both cases lead to a contradiction therefore

$Y_n \xrightarrow{d} Y \Rightarrow C_n \to Z$

Proof

Use of estimate of information in MLE CI

Definition

Mean Value Theorem

Proofish

Why $\bar{X}_n \pm Z_{\alpha/2} s/\sqrt{n}$ *works*

For the approximate large-sample CI for the MLE:

$$\sqrt{n\mathcal{I}(\hat{\theta}_n)}(\hat{\theta}_n - \theta) = \underbrace{\sqrt{n\mathcal{I}(\theta)}(\hat{\theta}_n - \theta)}_{\xrightarrow{d} N(0,1)} \underbrace{\sqrt{\frac{\mathcal{I}(\hat{\theta}_n)}{\mathcal{I}(\theta)}}}_{\xrightarrow{p} 1} \xrightarrow{d} N(0,1)$$

This is a consequence of asymptotic normality of the MLE, Slutsky's and CMT (because $\mathcal{I}(\hat{\theta}_n)$ is a continuous function of θ so if $\hat{\theta}_n \xrightarrow{p} \theta$ then $\mathcal{I}(\hat{\theta}_n) \xrightarrow{p} \mathcal{I}(\theta)$.)

$$\gamma(\hat{\theta}_n) = \gamma(\theta) + \gamma'(\widetilde{\theta})(\hat{\theta}_n - \theta)$$

This is helpful because you can rearrange to be:

$$\sqrt{n}(\gamma(\hat{\theta}_n) - \gamma(\theta)) = \gamma'(\widetilde{\theta})\sqrt{n}(\hat{\theta}_n - \theta)$$

Which allows you to show asymptotic normality of MLE

$$\frac{\sqrt{n}(\bar{X}_n - \theta)}{s} = \underbrace{\frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\sigma}{s}}_{\xrightarrow{p} 1} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

Pro Tip

When is the t-interval exact?

Statistical Inference

Proof

t-interval is robust to non-normality in large samples because...

Statistical Inference

Equation

Robust variance estimator of Normal Linear Regression

Statistical Inference

The t-interval is exact when $X_i \sim Normal$ because the pivot $\sqrt{n}(\bar{X}_n - \theta)/s$ is exactly t^{n-1}. It is approximately correct in large samples when the normality assumption fails because $t^{n-1}_{\alpha/2} \to Z_{\alpha/2}$ by the quantile convergence.

$$P\left(\frac{\sqrt{n}|\bar{X}_n - \theta|}{s} \leq t^{n-1}_{\alpha/2}\right) = P\left(\underbrace{\frac{\sqrt{n}|\bar{X}_n - \theta|}{\sigma}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\sigma}{s}}_{\xrightarrow{p} 1} \underbrace{\frac{Z_{\alpha/2}}{t^{n-1}_{\alpha/2}}}_{\xrightarrow{p} 1} \leq Z_{\alpha/2}\right) \to 1 - \alpha$$

$$\hat{\Lambda} = n(X'W^{-1}X)^{-1}X'W^{-1}diag\{r_i^2\}W^{-1}X(X'W^{-1}X)^{-1}$$

Weighted least squares is just least squares after scaling the data by the variance.

DEFINITION

Robust Large Sample Intervals

PRO TIP

Why does the robust large sample interval work?

DEFINITION

One parameter exponential family general case of robust large sample intervals

Basically, the same as before, just raise to the b/a in other words our new variance is just b/a^2. To estimate this:

$$\hat{\lambda} = \frac{n \sum \left(\frac{\partial l_i(\hat{\theta}_n)}{\partial \theta} \right)^2}{[\mathcal{I}_n(\hat{\theta}_n)]^2}$$

Here this $\mathcal{I}_n(\hat{\theta}_n)$ is the observed information: $-\sum \frac{\partial^2 \log f(y_i;\hat{\theta}_n)}{\partial \theta^2}$

$$P\left(\underbrace{\frac{\sqrt{n}|\hat{\theta}_n - \theta_0|}{\sqrt{\lambda}}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\sqrt{\lambda}}{\sqrt{\hat{\lambda}}}}_{\xrightarrow{p} 1} \leq Z_{\alpha/2} \right) \to 1 - \alpha$$

$$l_i(\theta; Y_i) = a(\theta) b(Y_i) + c(\theta)$$

$$\text{Where: } \sum l_i(\hat{\theta}_n; Y_i) = 0 \text{ and } \sum b(Y_i) = -n \frac{\partial c(\hat{\theta}_n)}{\partial \theta} \left(\frac{\partial a(\hat{\theta}_n)}{\partial \theta} \right)^{-1}$$

$$\text{Here: } \frac{n}{\mathcal{I}_n(\hat{\theta}_n)} = -\left\{ \frac{\partial^2 a(\hat{\theta}_n)}{\partial \theta^2} \frac{\sum b(Y_i)}{n} + \frac{\partial^2 c(\hat{\theta}_n)}{\partial \theta^2} \right\}^{-1}$$

$$\text{therefore: } \hat{\lambda} = \frac{n [a(\hat{\theta}_n)]^2 \sum \left[b(Y_i) - \frac{\sum b(Y_i)}{n} \right]^2}{[\mathcal{I}_n(\hat{\theta}_n)]^2}$$

Pro Tip

Weighted Least Squares Linear Regression if W misspecified

Statistical Inference

Pro Tip

If $\underline{Z} \sim MVN(\underline{\mu}, \Sigma)$ how is AZ distributed?

Statistical Inference

Pro Tip

If $\underline{Z} \sim MVN(\underline{\mu}, \Sigma)$ what would be distributed as χ^2_k?

Statistical Inference

$$Y \sim MVN(X, \underline{\beta}, \sigma^2 W) \text{ and } W = diag\{w_i\} = \begin{bmatrix} w_i & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_n \end{bmatrix}$$

The weighted least squares estimate of $\underline{\beta}$ is

$$\hat{\beta}_{wls} = X'W^{-1}X)^{-1}X'W^{-1}Y$$

This is a **consistent** estimator of $\underline{\beta}$ ever if the Ys are not normal and the covariance matrix is not proportional to W. BUT if W is misspecified that the variance-covariance matrix is not estimated by $n/\mathcal{I}$. So you need the robust variance estimator:

$$\hat{\Lambda} = n(X'W^{-1}X)^{-1}X'W^{-1}diag\{r_i^2\}W^{-1}X(X'W^{-1}X)^{-1}$$

$$AZ \sim MVN(A\underline{\mu}, A\Sigma A')$$

$$(\underline{Z} - \underline{\mu})'\Sigma^{-1}(\underline{Z} - \underline{\mu}) \sim \chi_k^2$$

This is only when Σ is full rank.

$$\Sigma = \sigma^2 \underline{\mathcal{I}}$$

$$(\underline{Z} - \underline{\mu})'\Sigma^{-1}(\underline{Z} - \underline{\mu}) = \Sigma(Z_i - \mu_i)^2/\sigma^2$$

$$f_{\underline{Z}}(Z) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(Z - \underline{\mu})'\Sigma^{-1}(Z - \underline{\mu})\right\}$$

Definition

Size of a test

Statistical Inference

Definition

Power

Statistical Inference

Pro Tip

General Setup for size and power

Statistical Inference

$$
\begin{aligned}
P(\text{Choose } H_1 \text{ when } H_0 \text{ is true}) &= P(\underline{\mathrm{X}} \in C_\delta | H_0) = P_0(\underline{\mathrm{X}} \in C_\delta) \\
&= P_0(\delta(\underline{\mathrm{X}}) = 1) \\
&= E_0[\delta(\underline{\mathrm{X}})] = E[\delta(\underline{\mathrm{X}})|H_0] \\
&= \int_{C_\delta} f(\underline{\mathrm{X}}; \theta_0) d\underline{\mathrm{X}} \\
&= \alpha
\end{aligned}
$$

$$
\begin{aligned}
1 - P(\text{Choose } H_0 \text{ when } H_1 \text{ is true}) &= P(\text{Choose } H_1 \text{ when } H_1 \text{ is true}) \\
&= P_1(\underline{\mathrm{X}} \in C_\delta) = P_1(\delta(\underline{\mathrm{X}}) = 1) \\
&= E_1[\delta(\underline{\mathrm{X}})] \\
&= \int_{C_\delta} f(\underline{\mathrm{X}}; \theta_1) d\underline{\mathrm{X}} \\
&= 1 - \beta
\end{aligned}
$$

$$
\begin{aligned}
X_1, ..., X_n &\overset{iid}{\sim} N(0,1) \\
H_0 : \theta = 0 &\text{ and } H_1 : \theta = 1 \\
\text{Test Stat: } \delta(\underline{\mathrm{X}}) &= \begin{cases} 1 & \bar{X}_n > c \\ 0 & \bar{X}_n \leq c \end{cases} \\
\text{Critical Region: } C_\delta &= \{\underline{\mathrm{X}} : \bar{X}_n > c\} \\
\alpha &= P_0(\bar{X}_n > c) = P(\sqrt{n}\bar{X}_n > \sqrt{n}c) = 1 - \Phi[\sqrt{n}c] \\
\beta &= P_1(\bar{X}_n \leq c) = P_1\left(\sqrt{n}\frac{(\bar{X}_n - \mu_1)}{\sigma} \leq \sqrt{n}\frac{(c - \mu_1)}{\sigma}\right) \\
&= \Phi\left[\sqrt{n}\frac{(c - \mu_1)}{\sigma}\right]
\end{aligned}
$$

Definition

Neyman-Pearson Lemma

Statistical Inference

Definition

Significance Testing

Statistical Inference

Definition

The Power Function

Statistical Inference

Using the LR will yield a most powerful test of size α

- Statistical procedure for measuring the strength of evidence against the null hypothesis - R.A. Fisher
- takes a test stat $T(\underline{X})$ where
  1. Larger values of $T(\underline{X})$ represents strong evidence of departure from H_0.
  2. Distribution of $T(\underline{X})$ under H_0 is known
  3. For given observations $\underline{x}$, the p-value is

$$p - value = P(T(\underline{X}) \geq T(\underline{x})|H_0)$$

- there are no rejection regions or alternative hypotheses.
- p-value are always in the tails of the null distribution
- answers "How do I interpret these observations as evidence"

The power function is the probability of rejecting H_0 (defined over $\Theta = \Theta_0 \cup \Theta_1$)

$$1 - \beta(\theta) = E_\theta[\delta(\underline{X})] \text{ for } \theta \in \Theta$$

Definition

Uniformly Most Powerful test

Definition

Unbiasedness of test

Definition

Composite Hypothesis Power Function

When $\delta(\underline{\mathrm{X}})$ is free of the alternative hypothesis, if N-P holds, then $\delta(\underline{\mathrm{X}})$ is UMP.

As long as the power $\geq$ size of the test, it is unbiased. In fancy speak: a size α test of $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$ is unbiased if

$$\inf_{\theta \in \Theta_1} 1 - \beta(\theta) \geq \alpha$$

UMP tests are unbiased. If a UMP test does not exist (like in 2-sided case) you can use UMPU - among unbiased tests, the UMP.

$$P_\theta(Reject\ H_0) = 1 - \beta(\theta) = E_\theta[\delta(\underline{\mathrm{X}})] \text{ for } \theta \in \Theta$$

Definition

Composite Hypothesis Size

Definition

Composite Hypothesis Consistency

Definition

Generalized Likelihood Ratio Test

$$\alpha = \sup_{\theta \in \Theta_0} 1 - \beta(\theta) = \sup_{\theta \in \Theta_0} E_\theta[\delta(\underline{\mathrm{X}})]$$

A series of tests $\delta_1, ..., \delta_n$ is consistent versus the alternative if $1-\beta_{\delta_n}(\theta)1 \to 1$ as $n \to \infty$

Reject H_0 if

$$\lambda(\underline{\mathrm{X}}) = \frac{\sup_{\theta \in \Theta_0} f(\underline{\mathrm{X}}; \theta)}{\sup_{\theta \in \Theta} f(\underline{\mathrm{X}}); \theta)} = \inf_{\theta \in \Theta} \sup_{\theta \in \Theta_0} f(\underline{\mathrm{X}}; \theta)$$

is too small.

Specifically, reject $H_0 : \theta \in \Theta_0$ if $\lambda(\underline{\mathrm{X}}) \leq \lambda_0$ where $\alpha = \sup_{\theta \in \Theta_0} P_\theta(\lambda(\underline{\mathrm{X}}) \leq \lambda_0)$.

Pro Tip

Interpreting GLRT

Statistical Inference

Definition

Monotone Likelihood Ratio Property

Statistical Inference

Definition

Karlin-Rubin Theorem

Statistical Inference

Do not interpret test as "evidence" for or against a composite hypothesis. It is just saying you can find one simple alternative that is better supported than each null hypothesis. It does not mean that the alternative as a set is better than the set of null hypothesis.

A family of pdfs or pmfs with univariable random variable t and a parameter θ has a MLR if

$\forall\, \theta_2 > \theta_1$ we have that $g(t|\theta_2)/g(t|\theta_1)$ is a monotone ($\uparrow$ or $\downarrow$) function of t.

Let $T(\underline{\mathrm{X}})$ be a sufficient statistic for θ. If $\{g(t|\theta); \theta \in \Theta\}$ has the MLR property, then for any t_0 the test of $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ rejects if $T(\underline{\mathrm{X}})$ is a UMP test of size $\alpha = P_{\theta_0}(T(\underline{\mathrm{X}}) > t_0)$

Pro Tip

Exponential Family MLRs

Pro Tip

GLRT distribution

Definition

Wald Test

In the exponential family, we have $h(\underline{X})c(\theta)\exp\{w(\theta)T(\underline{X})\}$. If $w(\theta)$ is increasing, by Karlin-Rubin test:

$$\delta(\underline{X}) = \begin{cases} 1 & T(\underline{X}) > t_0 \\ 0 & T(\underline{X}) < t_0 \\ \gamma & T(\underline{X}) = t_0 \end{cases}$$

is UMP with size $\alpha = P_{\theta_0}(T(\underline{X}) > t_0)$ for testing $(\theta_1 > \theta_0)$

$$\frac{f(\underline{X};\theta_1)}{f(\underline{X};\theta_0)} > k \iff \frac{c(\theta_1)\exp\{w(\theta_1)T(\underline{X})\}}{c(\theta_0)\exp\{w(\theta_0)T(\underline{X})\}} > k \iff T(\underline{X})[w(\theta_1) - w(\theta_0)] > k^*$$

In a large sample...

$$-2\log\lambda(\underline{X}) \xrightarrow{d} \chi^2_{d-d_0}$$

where $\dim\Theta = d$ and $\dim\theta_0 = d_0$

Wald tests are based on an estimate of the information that is consistent under either the null or alternative hypothesis. A Wald test for the MLE is based on the asymptotic normality of the MLE.

$$\frac{(\hat{\theta}_n - \theta_0)}{\sqrt{Var(\hat{\theta}_n)}} \sim N(0,1)$$
$$\frac{(\hat{\theta}_n - \theta_0)^2}{Var(\hat{\theta}_n)} \sim \chi^2_1$$

Not invariant to transformations of the parameter space. Do not work well when the true parameter is near the edge of the parameter space because the normal approximation often fails.

Definition

Score test statistic

Statistical Inference

Definition

How is regression GLRT distributed when using $\hat{\sigma}$ instead of σ

Statistical Inference

Definition

FDR from 2x2 table

Statistical Inference

$$\frac{S_n(\theta_0)}{\sqrt{\mathcal{I}_n(\theta_0)}} \sim N(0,1)$$
$$\frac{[S_n(\theta_0)]^2}{\mathcal{I}_n(\theta_0)} \sim \chi_1^2$$

This holds ins large samples under H_0 because of AN of score function. Notice the information is calculated under the null hypothesis.
Examples:

- Cochran-Mantel-Haenzel test
- Log-Rank test

Most powerful test for ‘small’ deviations from H_0 by N-P lemma.

You can estimate σ^2 with $\hat{\sigma}^2$

$$-2\log\lambda(X) \approx \frac{[RSS(x_1) - RSS(x_1,x_2)]/(p_l - p_s)}{RSS(x_1,x_2)/(n-p_l)} \sim F_{p_l-p_s,n-p_l}$$

where p_i is the number of parameters in model i

Note $F_{1,n-3} \to \chi_1^2$ as $n \to \infty$

	H_0 Accepted	H_0 Rejected	Total
H_0 True	U	V	M_0
H_0 False	T	S	M_1
Total	M-R	R	M

$FDP_0 = V/R$
$FDR = E[FDP_0] = E\left[\frac{V}{R}|R>0\right]P(R>0)$

Definition

Bootstrap

Statistical Inference

Definition

Law of the Iterated Logarithm

Statistical Inference

Definition

Marginal Likelihood

Statistical Inference

The bootstrap algorithm:

1. Draw bootstrap sample
2. Compute a statistic
3. Repeat B times to get a new statistic each time
4. Compute:
$$V_{Boot} = \frac{1}{B}\left(T^*_{n,i} - \frac{1}{n}\sum T^*_{n,j}\right)^2$$
5. Use V_{Boot} as an approximation to the variance of the test statistic
6. Generate CI using: Normal Interval, Percentile Interval, Pivotal Interval, Bias-Corrected Interval

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n\log\log n}} = \sqrt{2} \text{ almost surely}$$

Here limsup is the limiting supremum.

$$L_m(\theta;\underline{x}) = f(\underline{x};\theta) = \int_\Gamma f(\underline{x};\theta,\gamma)g(\gamma)d\gamma$$

$g(\gamma)$ is the pdf

Definition

Estimated Likelihood

Statistical Inference

Definition

Profile Likelihood

Statistical Inference

Definition

Conditional Likelihood

Statistical Inference

$$L_e(\theta; x) = L_e(\theta, \hat{y} l \underline{\mathrm{x}})$$

where $\hat{\gamma}$ is an estimator of γ based on the data. This is just a fixed estimate, versus the profile likelihood which finds the best guess conditional on the parameter value.

$$L_p(\theta; \underline{\mathrm{x}}) = \sup_{\gamma \in \Gamma} L(\theta, \gamma; \underline{\mathrm{x}}) = L(\theta, \hat{\gamma}(\theta); \underline{\mathrm{x}})$$

$\hat{\gamma}(\theta)$ is the MLE of γ given θ.

$$L_c(\theta; \underline{\mathrm{x}}) = f(\underline{\mathrm{x}} | \theta, S(\underline{\mathrm{x}}))$$

$S(\underline{\mathrm{x}})$ is a special statistic and conditioning on it makes the likelihood free of γ.

PRO TIP

Quasi-log-likelihood regression example

DEFINITION

Limiting Expectation

DEFINITION

Asymptotic Expectation

$$l(\beta|\underline{\mathrm{y}}, \underline{\mathrm{z}}) = \frac{1}{\sigma^2} \sum y_i(\beta_0 + \beta_1 z_i) - e^{(\beta_0+\beta_1 z_i)} - \log y_i!$$

$$\lim_{n\to\infty} E[X_n]$$

$$E[\lim_{n\to\infty} X_n]$$

Definition

Bayes Factor

Statistical Inference

Definition

Confidence Interval with Indifference Zone

Statistical Inference

Pro Tip

When is t-interval CI valid

Statistical Inference

A likelihood ratio that is a ratio of marginal likelihoods is called a Bayes factor:

$$BF_{0,1}\frac{P(\underline{\mathrm{X}}|H_0)}{P(\underline{\mathrm{X}}|H_1)} = \frac{\int_{\Theta_0} f(x;\theta)g_0(\theta;\gamma_0)d\theta}{\int_{\Theta_1} f(x;\theta)g_1(\theta;\gamma_1)d\theta} = \frac{f(X;\theta_0)}{f(X;\theta_1)} = \frac{L_n(\theta_0)}{L_n(\theta_1)}$$

$$BF_{0,1} = \frac{P(H_0|X=x)P(H_1)}{P(H_1|X=x)P(H_0)}$$

$$P(\Delta \cap I(\underline{\mathrm{X}}) = \emptyset) = 2\Phi[-(\delta\sqrt{n} + Z_{\alpha/2})]$$

This is good because

$$P(\Delta \cap I(\underline{\mathrm{X}}) = \emptyset) \to 0 \text{ as } n \to \infty \ \forall \delta \neq 0$$

and

$$P(\Delta \cap I(\underline{\mathrm{X}}) = \emptyset) = \alpha \text{ when } \delta = 0$$

$$\bar{X}_n - \bar{Y}_m \pm t_{\alpha/2}^{n+m-2}\sqrt{S_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}$$

This CI is valid - in other words has **exactly** $100(1-\alpha)\%$ coverage probability when

1. The population distribution of each group is normal
2. The groups have a common variance

However in a large sample, the equal-variance t-interval is a valid approximate large sample CI that is robust to non-normality. Either the variances need to be equal OR the sample size of the groups needs to be equal.

Definition

Randomized Test

Statistical Inference

Definition

N-P Randomized Test

Statistical Inference

Definition

Strength of Evidence from a Hypothesis test

Statistical Inference

A randomized test randomly chooses between the competing hypotheses in certain situations:

$$\delta(\underline{\mathrm{X}}) = \begin{cases} 1 & \text{Choose } H_1 \\ 0 & \text{Choose } H_0 \\ \gamma & \text{Choose } H_1 \text{with probability } \gamma \end{cases}$$

This is generally done to increase power.

N-P LRT:

$$\delta(\underline{\mathrm{X}}) = \begin{cases} 1 \text{ (Choose } H_1) & \frac{f_1(X)}{f_0(X)} > k \\ \gamma \text{ (Choose ?)} & \frac{f_1(X)}{f_0(X)} = k \\ 0 \text{ (Choose } H_0) & \frac{f_1(X)}{f_0(X)} < k \end{cases}$$

By randomizing on the decision boundary we can increase power and maintain a test of certain size

Support of H_1 over H_0 by the factor:

$$\frac{P(\delta(\underline{\mathrm{x}}) = 1|H_1)}{P(\delta(\underline{\mathrm{x}}) = 1|H_0)} = \frac{1-\beta}{\alpha}$$

Pro Tip

If the outcome only of the test is reported, what size study provides more evidence in support of H_1 over H_0?

Statistical Inference

Pro Tip

If the p-value is reported, what size study provides more evidence in support of H_1 over H_0?

Statistical Inference

Pro Tip

Post-hoc power calculations

Statistical Inference

The larger study provides more evidence

The smaller study provides more evidence

The observed power of a test is a simple 1-to-1 function of the p-value.

Definition

What is the collection of null hypotheses that fail to reject at the α level?

Statistical Inference

Definition

Bounds on Bayes Factor

Statistical Inference

Pro Tip

If $T \sim t_{\alpha/2}^{n+m-k}$ how is T^2 distributed?

Statistical Inference

The $100(1-\alpha)\%$ confidence interval!

$$\frac{\inf_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta_1} L(\theta)} \leq BF_{0,1} \leq \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\inf_{\theta\in\Theta_1} L(\theta)}$$

$$T^2 \sim F_{k-2,n-k}$$

Pro Tip

F-test robustness

Statistical Inference

Pro Tip

Working Model is Correct

Statistical Inference

Pro Tip

Working Model is Incorrect

Statistical Inference

The F-test for means is robust to departures from normality (but sensitive to equal variance assumption or equal n in each group), BUT the F-test for variances is not! LALALa.

- $\hat{\theta}_n$ (the MLE of θ) is consistent for θ_0 as $n \to \infty$
- As $n \to \infty$ $\frac{L_n(\theta)}{L_n(\theta_0)} \xrightarrow{a.s.} 0$
- The probability of observing misleading evidence:

$$P\left(\frac{L_n(\theta)}{L_n(\theta_0)} \geq k\right) \to 0$$
$$P\left(\frac{L_n(\theta_n)}{L_n(\theta_0)} \geq k\right) \to \Phi\left[-\frac{\log k}{|c|} - \frac{|c|}{2}\right]$$

Where: $\theta_n = \theta_0 + c/\sqrt{n}$

- The asymptotic behavior of the likelihood ratio is:

$$\log\left\{\frac{L_n(\theta_n)}{L_n(\theta_0)}\right\} \xrightarrow{d} N\left(\frac{-c^2}{2}, c^2\right)$$

- The maximum probability of observing misleading evidence is:

$$\max P\left(\frac{L_n(\theta)}{L_n(\theta_0)} \geq k\right) \to \Phi\left[-\sqrt{2\log k}\right]$$

- $\hat{\theta}_n$ is the MLE and $\hat{\theta}_n \to \theta_g$ as $n \to \infty$ and $\theta_g = argmax E_g[\log f(x_i;\theta)]$
- As $n \to \infty$: $\frac{L_n(\theta)}{L_n(\theta_g)} \xrightarrow{a.s} 0$
- The probability of observing misleading evidence:

$$P\left(\frac{L_n(\theta)}{L_n(\theta_0)} \geq k\right) \to 0$$
$$P\left(\frac{L_n(\theta_n)}{L_n(\theta_0)} \geq k\right) \to \Phi\left[-\frac{\log k}{|c|\sqrt{b}} - \frac{|c|a}{2\sqrt{b}}\right] \text{Where: } \theta_n = \theta_0 + c/\sqrt{n}$$

- The asymptotic behavior of the likelihood ratio is:

$$\log\left\{\frac{L_n(\theta_n)}{L_n(\theta_0)}\right\} \xrightarrow{d} N\left(\frac{-c^2 a}{2}, c^2 b\right)$$

- The maximum probability of observing misleading evidence is:

$$\max P\left(\frac{L_n(\theta)}{L_n(\theta_0)} \geq k\right) \to \Phi\left[-\sqrt{2\log k}\sqrt{\frac{a}{b}}\right]$$

Pro Tip

Under the robust working model

Statistical Inference

Pro Tip

Doing Quasi-likelihood

Statistical Inference

Pro Tip

Trick to show $MSE = Var[\hat{\theta}] + b^2(\hat{\theta})$

Statistical Inference

- $\hat{\theta}_n$ is the MLE and $\hat{\theta}_n \to \theta_g$ as $n \to \infty$ and $\theta_g = argmax E_g[\log f(x_i;\theta)]$
- As $n \to \infty$: $\frac{L_n(\theta)}{L_n(\theta_g)} \xrightarrow{a.s} 0$
- The probability of observing misleading evidence:

$$P\left(\frac{L_n(\theta)}{L_n(\theta_0)} \geq k\right) \to 0$$

$$P\left(\frac{L_n(\theta_n)}{L_n(\theta_0)} \geq k\right) \to \Phi\left[-\frac{\log k\sqrt{b}}{|c|a} - \frac{|c|a}{2\sqrt{b}}\right] \text{ Where: } \theta_n = \theta_0 + c/\sqrt{n}$$

- The asymptotic behavior of the likelihood ratio is:

$$\frac{\hat{a}}{\hat{b}} \log\left\{\frac{L_n(\theta_n)}{L_n(\theta_0)}\right\} \xrightarrow{d} N\left(\frac{-c^2a^2}{2b}, \frac{c^2a^2}{b}\right)$$

- The maximum probability of observing misleading evidence is:

$$\max P\left(\frac{L_n(\theta)}{L_n(\theta_0)} \geq k\right) \to \Phi\left[-\sqrt{2\log k}\right]$$

---

1. Start with the score-function:

$$\sum \frac{Y_i - E[Y_i]}{Var[Y_i]}$$

2. Set score function equal to 0 and get solve for θ to get a quasi-MLE
3. Then you also get a quasi-information in order to better estimate the variance
4. If you want to get the quasi log likelihood, you integrate the quasi-score function

---

JUST ADD AND SUBTRACT $E[\hat{\theta}]$ WHAT? yeah. that was a comp question. PLEASE GIVE US THAT AGIAN!

$$\begin{aligned} E[(\hat{\theta}_n - \theta)^2] &= E[(\hat{\theta}_n - E[\hat{\theta}] + E[\hat{\theta}] - \theta)2] \\ &= E[(\hat{\theta}_n - E[\hat{\theta}_n])^2] + (E[\hat{\theta}] - \theta)^2 \\ &= Var[\hat{\theta}] + b^2(\hat{\theta}) \end{aligned}$$

Pro Tip

Variance of X^2 in Normal Distribution

Statistical Inference

Pro Tip

Fisher's Information for Normal

Statistical Inference

Pro Tip

Find the median of a distribution

Statistical Inference

$$
\begin{aligned}
Var(X_2) &= E[X^4] - (E[X^2])^2 \\
&= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^2 - (\sigma^2 + \mu^2)^2 \\
&= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^2 - \sigma^4 - 2\sigma^2\mu^2 - \mu^4 \\
&= 4\mu^2\sigma^2 + 3\sigma^2 - \sigma^4
\end{aligned}
$$

$$
\begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}
$$

Set the cdf$=\frac{1}{2}$

Pro Tip

Location Scale MGF

Statistical Inference

Pro Tip

$Logit(\beta_0 + \beta_1 X)$

Statistical Inference

Pro Tip

Probit Model

Statistical Inference

$$Mgf_x(t) = Mgf_{\sigma z+\mu}(t) = e^{\mu t} Mgf_z(\sigma t)$$

$$\Lambda(\beta_0 + \beta_1 X) = \frac{\exp\{\beta_0 + \beta_1 X\}}{1 + \exp\{\beta_0 + \beta_1 X\}}$$

$$P(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$$

Pro Tip

Show Asymptotic Bias

Statistical Inference

Pro Tip

Exact Binomial Distribution, McNemars

Statistical Inference

Proof

Goodness of Fit Classic Chi-Square Formula

Statistical Inference

Arrange what you are interested in to see where it goes in the limit (multiply by $\sqrt{n}$) If there is any bias, then that is the asymptotic bias.

$$Binomial(b+c, 0.5) \to N((b+c)(0.5), (b+c)(0.5)^2)$$

Then compare this to b:

$$\frac{b - \left(\frac{b+c}{2}\right)}{\frac{\sqrt{b+c}}{2}} \sim N(0,1)$$
$$\frac{b-c}{\sqrt{b+c}} \sim N(0,1)$$
$$\frac{(b-c)^2}{b+c} \sim \chi_1^2$$

$$\underline{\mathrm{X}} \sim Mult(n, \underline{\theta}) \quad \lambda(\underline{\mathrm{X}}) = \prod_{i=1}^{m} \left(\frac{\gamma_i}{\hat{\theta}_i}\right)^{x_i}$$

$n\gamma_i$=expected cell count and $X_i = n\hat{\theta}_i$

$$-2\log(\lambda(\underline{\mathrm{X}})) = -2\sum n\hat{\theta}_i \log\left(\frac{\gamma_i}{\hat{\theta}_i}\right)$$

By Taylor Series Expansion: $x\log\left(\frac{x}{x_0}\right) \approx (x - x_0) + \frac{(x-x_0)^2}{2x_0}...$

$$= n2\sum(\hat{\theta}_i - \gamma_i) + \frac{2n\sum(\hat{\theta}_i - \gamma_i)^2}{2\gamma_i}$$
$$\sum \hat{\theta}_i = \sum \gamma_i = 1$$
$$= \sum_{i=1}^{m} \frac{(x_i - n\gamma_i)^2}{n\gamma_i}$$

Pro Tip

$E(z_1|z_2)$ when (z_1, z_2) is bivariate Normal

Statistical Inference

Pro Tip

Posterior Normal Distribution

Statistical Inference

Pro Tip

How can you split up $(1-p)^w$

Statistical Inference

$$E(z_1|z_2) = E(z_1) + \frac{Cov(z_1, z_2)}{Var(z_2)}(z_2 - E(z_2))$$
$$Var(z_1|z_2) = Var(z_1) - \frac{Cov^2(z_1, z_2)}{Var(z_2)}$$

$$\mu|x_1, ..., x_n \sim N\left(\frac{\sigma_0^2}{\sigma^2 + n\sigma_0^2}\bar{x} + \frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\mu_0, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$$

$$(1-p)^x(1-p)^{w-x}$$

This is sometimes useful if you are trying to recreate a binomial distribution and you have $(something)^w$

Proof

Prove the accept-reject algorithm

Proof

Maximizing the likelihood under model failure

Proof

Prove the Memoryless Property of Exponential

$$
\begin{aligned}
P(Y \leq y) &= P\left(V \leq y | U \leq \frac{1}{M}\frac{f_y(v)}{f_v(v)}\right) \\
&= \frac{P\left(V \leq y, U \leq \frac{1}{M}\frac{f_y(v)}{f_v(v)}\right)}{P\left(U \leq \frac{1}{M}\frac{f_y(v)}{f_v(v)}\right)} \\
&= \frac{\int_{-\infty}^{y}\int_0^{\frac{1}{M}\frac{f_y(v)}{f_v(y)}} 1du f_v(v)dv}{\int_{-\infty}^{\infty}\int_0^{\frac{1}{M}\frac{f_y(v)}{f_v(v)}} 1du f_v(v)dv} \\
&= \frac{\int_{-\infty}^{y} \frac{1}{M}\frac{f_y(v)}{f_v(v)} f_v(v)dv}{\int_{-\infty}^{\infty} \frac{1}{M}\frac{f_y(v)}{f_v(v)} f_v(v)dv} \\
&= \frac{\int_{-\infty}^{y} f_y(v)dv}{\int_{-\infty}^{\infty} f_y(v)dv} = \frac{F_Y(y)}{1} = F_Y(y)
\end{aligned}
$$

$$
\begin{aligned}
\hat{\theta}_n &= \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_i \log f(X_i;\theta) \\
&= \underset{\theta\in\Theta}{\operatorname{argmin}} \frac{1}{n}\sum_i \log \frac{1}{f(X_i;\theta)} \\
&= \underset{\theta\in\Theta}{\operatorname{argmin}} \frac{1}{n}\sum_i \log \frac{1}{f(X_i;\theta)} + \frac{1}{n}\log \hat{g}(x_i) \\
&= \underset{\theta\in\Theta}{\operatorname{argmin}} \frac{1}{n}\sum_i \log \frac{\hat{g}(x_i)}{f(X_i;\theta)}
\end{aligned}
$$

Assume $t > s$

$$
\begin{aligned}
P(X > t | X > s) &= P(X > t - s) \\
P(X > t | X > s) &= \frac{P(X > t, X > s)}{P(X > s)} \\
&= \frac{P(X > t)}{P(X > s)} = \frac{1 - F(t)}{1 - F(s)} \\
&= \frac{1 - (1 - e^{-t\beta})}{1 - (1 - e^{-s\beta})} \\
&= e^{-\beta(t-s)} \\
&= 1 - F(t - s) \\
&= P(X > t - s)
\end{aligned}
$$

Proof

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

Statistical Inference

Pro Tip

Beta $E[X^n]$

Statistical Inference

Proof

Central Limit Theorem

Statistical Inference

$$\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{U}=\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2}_{V}+\underbrace{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}_{W}$$

$$U\sim\chi^2(n)\quad W\sim\chi^2(1)$$

$$V=\frac{(n-1)S^2}{\sigma^2}$$

$tV\perp W$ because $\bar{X}\perp S^2$ therefore:

$$M_u(t)=M_V(t)M_W(t)$$

$$\frac{1}{(1-2t)^{n/2}}=M_V(t)\frac{1}{(1-2t)^{1/2}}$$

$$M_V(t)=\frac{1}{(1-2t)^{(n-1)/2}}$$

$$V\sim\chi^2(n-1)$$

$$E[X^n]=\frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}$$

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}=\frac{1}{\sqrt{n}}\sum Z_i\quad M_{\frac{1}{\sqrt{n}}\sum Z_i}(t)=\left(M_y\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Taylor series: $M_z\left(\frac{t}{\sqrt{n}}\right)=M_z(0)+M_z'(0)\left(\frac{t}{\sqrt{n}}-0\right)+\frac{M_z''(0)\left(\frac{t}{\sqrt{n}}-0\right)^2}{2!}+\ldots$

$$M_z'(0)=E(Z)=E\left(\frac{x-\mu}{\sigma}\right)=\frac{E(X)-\mu}{\sigma}=0$$

$$M_z''(0)=Var(Z)+(E(Z))^2=Var\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sigma^2}Var(X)=\frac{\sigma^2}{\sigma^2}=1$$

$$M_z(0)=E(e^{0z})=1$$

$$M_z\left(\frac{t}{\sqrt{n}}\right)=1+0+\frac{t^2}{2n}\Rightarrow\left(M_z\left(\frac{t}{\sqrt{n}}\right)\right)^n\approx\left(1+\frac{\frac{t^2}{2}}{n}\right)^n$$

$$\Rightarrow\lim_{n\to\infty}\left(1+\frac{\frac{t^2}{2}}{n}\right)^n=e^{t^2/2}$$

Proof

Kullback-Leibler Divergence

Statistical Inference

Pro Tip

If $E(Y|X)$ is linear then how can you write this in terms of μ_x, μ_y, ρ_{xy}, σ_x, and σ_y

Statistical Inference

Pro Tip

Inverse Gamma - Normal Posterior

Statistical Inference

$$
\begin{aligned}
E_g\left[\log\frac{g(x)}{f(x)}\right] &= E_g\left[-\log\frac{f(x)}{g(x)}\right]\\
E_g\left[-\log\frac{f(x)}{g(x)}\right] &\geq -\log E_g\left[\frac{f(x)}{g(x)}\right]\\
&\geq -\log(1)\\
&\geq 0
\end{aligned}
$$

$$
\begin{aligned}
E(Y|X=x) &= a + bx\\
&= \mu_y - \rho_{xy}\frac{\sigma_y}{\sigma_x}\mu_x + \rho_{xy}\frac{\sigma_y}{\sigma_x}x
\end{aligned}
$$

Normal model with known mean, unknown variance.

$$
\begin{aligned}
\sigma^2 &\sim IG(\alpha_0, \beta_0)\\
\underline{\mathrm{X}} &\sim N(\mu, \sigma^2)\\
Posterior &= IG\left(\alpha_0 + \frac{n}{2}, \beta_0 + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right)
\end{aligned}
$$

Proof

Poisson MGF

Statistical Inference

Pro Tip

What is $e^{-\lambda}$ in terms of summation

Statistical Inference

Pro Tip

How to get variance of of MOM

Statistical Inference

$$
\begin{aligned}
E[e^{tx}] &= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} \\
&= e^{-\lambda + \lambda e^t} \\
&= e^{\lambda(e^t - 1)}
\end{aligned}
$$

$$
\begin{aligned}
1 &= \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \\
e^{-\lambda} &= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}
\end{aligned}
$$

DELTA METHOD

Pro Tip

Credible Interval

Statistical Inference

Pro Tip

Geometric Series

Statistical Inference

Pro Tip

Geometric MFG

Statistical Inference

$$LB = F^{-1}_{\alpha/2}$$
$$UB = F^{-1}_{1-\alpha/2}$$

$$\sum_{x=1}^{\infty} r^x = \frac{r}{1-r}$$
$$\sum_{x=0}^{\infty} r^x = \frac{1}{1-r}$$

$$\frac{pe^t}{1-(1-p)e^t}$$

Test

McNemar's Test

Test

Wilcoxon Sign Rank Test

Confidence Interval

Relative Risk

McNemar's uses the normal approximation to the Binomial. This requires large sample to be valid.

$$\chi_1^2 = \frac{(b-c)^2}{b+c}$$

To get a p-value:

$$b \sim Binomial(b+c, .5)$$
$$p - value = P(b > b_{obs} | n = b + c, \theta = 1/2)$$
$$P(X \geq x | n, \theta) = \sum_{i=x}^{n} \binom{n}{i} .5^n$$

two sided would be this p-value multiplies by 2.

Tests difference in distribution. For a one sided test, the test statistic is the sum of the ranks. For example if you are testing if group 2 is greater than group 1, then you order all of the estimates and rank them and then take out all of group 2's ranks and add them together!

$$p - value = P(W_2 \geq w_1)$$

CI for Relative Risk:

$$\exp\left\{\log\left(\frac{a/(a+b)}{c/(c+d)}\right) \pm Z_{\alpha/2}\sqrt{\frac{b/a}{b+a} + \frac{d/c}{d+c}}\right\}$$

Confidence Interval

Risk Difference

Confidence Interval

Odds Ratio

Pro Tip

Semi-partial correlation

CI for Risk Difference:

$$p_1 = \frac{a}{a+b}$$
$$n_1 = a + b$$
$$p_2 = \frac{c}{c+d}$$
$$n_2 = c + d$$

$$p_1 - p_2 \pm Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

CI for Odds Ratio:

$$\log\left(\frac{ad}{bc}\right) \pm Z_{\alpha/2}\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Regress X_1 with all the other covariates

$$X_1 \sim X_2 + X_3 + X_4$$

get the errors, e_1

Get the correlation of e_1 and Y

Pro Tip

Partial correlation

Definition

95% *Confidence Interval*

Definition

95% *Support Interval*

Regress X_1 with all the other covariates

$$X_1 \sim X_2 + X_3 + X_4$$

get the errors, e_1

Regress Y with all the remaining covariates

$$Y \sim X_2 + X_3 + X_4 \text{ get the errors, } e_Y$$

Get the correlation of e_1 and e_Y

The procedure will capture the true parameter 95% of the time. The values come from a procedure that tends to capture the true value μ

This is the interval that the data supports at the $1/k$ level.

Definition

95% Credible Interval

Assumptions

Logistic Regression

Interpretation

β*s a Logistic Regression*

μ is treated as random, the probability that μ is within these intervals is 95%

- The outcome needs to be binary
- Correct specification of the model
- The error terms need to be independent

Let's say we have a model $logit(p) = \beta_0 + \beta_1 * female$
β_1 is the log odds ratio for the between the female group and the male group.
β_0 is the log odds for males.
If it were a continuous model, say $logit(p) = \beta_0 + \beta_1 * testscore$
Then β_1 would be the difference in the log odds, in other words, for a one-unit increase in test score, the expected change in log odds is β_1. If you exponentiate it, then e^{β_1} is the odds. So for example, If it was 1.18, then you would expect to see an 18% increase in odds of being in group 1.
If there are two predictors, $logit(p) = \beta_0 + \beta_1 * female + \beta_2 * testscore$
Then holding test score constant, the odds of Y=1 for females over the odds of Y=1 for males is e^{β_1}.
β_2 is Holding female constant we will see a $X\%$ $(1 - e^{\beta_2})$ increase in the odds of Y=1 for a one-unit increase in test score.

Interpretation

β in a Linear Regression

Pro Tip

Why would you want to center variables in regression

Equation

SXX, sum of squares for the xs

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

- β_0 is the value you would predict for Y when all Xs are at the reference group.

- β_1 is the difference in the predicted value of Y for each one-unit difference in X1 if X2 remains constant.

- Centering variables so that the predictors have a mean of 0 make the intercept term interpreted as the expected value of Y_i when the predictor values are set to their means. Otherwise the intercept is interpreted as the expected value of Y_i when the predictors are set to 0, which may not be realistic.
- The sample covariance matrix of a matrix of values centered by their sample means is just X'X.
- If a univariate random variable is mean centered, then $var(X) = E(X^2)$ and the variance can be estimated from a sample by looking at the sample mean of squares of the observed values

$$\sum(x_i - \bar{x})^2$$

Equation

SD_x^2 *Sample variance of xs*

Equation

SXY, sum of the cross-products

Equation

s_{xy} *Sample covariance*

$SXX/(n-1)$

$\sum(x_i - \bar{x})(y_i - \bar{y})$

$SXY/(n-1)$

Equation

r_{xy} *Sample correlation*

Equation

$\hat{e}_i$, *the residual*

Equation

$\hat{\beta}_1$

$s_{xy}/(SD_x SD_y)$

$$\begin{aligned}
\hat{e}_i &= y_i - \hat{E}(Y|X = x_i)\\
&= y_i - \hat{y}\\
&= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_1 &= \frac{SXY}{SXX}\\
&= r_{xy}\frac{SD_y}{SD_x}\\
&= r_{xy}\left(\frac{SYY}{SXX}\right)^{1/2}
\end{aligned}$$

Equation

$\hat{\beta}_0$

Equation

RSS, residual sum of squares

Equation

R^2

$$\bar{y} - \hat{\beta}_1 \bar{x}$$

$$RSS = SYY - \hat{\beta}_1^2 SXX$$

$$\begin{aligned}
R^2 &= \frac{SSReg}{SYY} \\
&= \frac{(SXY)^2}{SXXSYY} \\
&= r_{xy} \\
SSReg &= SSY - RSS
\end{aligned}$$

Equation

χ^2 Expected cells

Equation

Chi-square test statistic

Equation

95% CI

$$\frac{\textit{row sum} \times \textit{column sum}}{N}$$

$$= \frac{(a+b)(a+c)}{a+b+c+d} \text{ for example for the top left cell}$$

$$\chi^2_{(r-1)(c-1)} = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

$$\bar{X}_n \pm Z_{\alpha/2} \hat{SE}(\bar{X}_n)$$

For one sided, reject if:

$$\frac{\bar{X}_n - \mu_0}{\hat{SE}(\bar{X}_n)} > 1.64$$

Definition

Mantel-Haenszel test

Definition

Fisher's Exact test

Pro Tip

Interpret Root MSE

This is an alternative to pooling data from multiple 2x2 tables - it helps avoid Simpson's paradox.

- the odds ratio for the ith strata is: $OR_i = \frac{a_i d_i}{b_i c_i}$
- The summary OR is $OR_{pooled} = \frac{\sum a_i \sum d_i}{\sum b_i \sum c_i}$
- The MH is $O_{MH} = \sum w_i OR_i / \sum w_i$ where $w_i = b_i c_i / N$
- $H_0 : OR_{MH} = 1$ test statistic

$$\frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{n}{a+b}}$$

- The sample standard deviation of the differences between predicted and observed values.
- Measure of accuracy

Pro Tip

Interpret R^2

Principles of Modern Biostatistics

- The amount of variability the model explains of the response data around its mean
- How close the data are to the fitted regression
- $\frac{SXY^2}{SXXSYY}$