Statistics

Steven Wang

2016

Contents

1 Probability

-Combination:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

-Permutation:

$$P_k^n = \frac{n!}{(n-k)!}$$

-Conditional Probability:

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

2 Distributions

-Binomial: $x \sim B(n,p)$

PMF:

$$f(x|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E[x] = np, Var[x] = np(1-p)$$

-Normal:

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

-Exponential:

PDF:

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

λ is # of events per unit of time.

CDF:

$$F(x|\lambda) = 1 - e^{-\lambda x}$$

$$E[x] = \lambda^{-1}, Var(x) = \lambda^{-2}$$

-Poisson:

PMF:

$$P = \frac{\lambda^k e^{-\lambda}}{k!}$$

k events in interval

$$E[x] = \lambda, Var[x] = \lambda$$

-Chi-Squared
-Student-t
-F-distribution

-Multivariate normal distribution:

$$MVN(\mu, \Sigma) = f(x_1, x_2, ..., x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu))$$

$$T\Sigma_{MLE} \sim W(\Sigma, T-1, m)$$

(Wishart Distribution)

3 Expectation and Mean Value

$$E(e^x) = e^{E(x) + \frac{1}{2}Var(x)}$$

Proof by using the definition of the probability density function of Normal Distribution (integrate the pdf)

4 Variance and Covariance

$$Var(X) = E[(X-\mu)^2] = E[x^2] - E[x]^2$$
$$Cov(X,Y) = \sigma_{X,Y} = E[(X-E[X])(Y-E[Y])] = E(XY) - E[X]E[Y]$$
$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X, \sigma_Y}$$

$$Var(X) = E_Y(Var(X|Y)) + Var[E(X|Y)]$$

5 Hypothesis testing

$\alpha =$ Type I error $=$ Prob(Reject $H_0|H_0$ is True)
$\beta =$ Type II error $= 1 - Power =$ 1 - Prob(Accept $H_0|H_0$ is Wrong)

If large sample, reduce α. From small sample size to large sample size:
10% * 5% ** 1% ***

$$Power = P(\frac{\hat{\beta}-0}{SE} > t_{0.975}|\beta = \beta_1)$$
$$= P(\frac{\hat{\beta}-\beta_1+\beta_1-0}{SE} > t_{0.975}|\beta = \beta_1) + P(\frac{\hat{\beta}-\beta_1+\beta_1-0}{SE} < t_{0.025}|\beta = \beta_1)$$

z-score(z-Distribution):
$$z = \frac{x-\mu}{\sigma}$$

t-statistic(t-Distribution):

$$t = \frac{\hat{x}-x_0}{s.e.(\hat{x})} = \frac{\hat{x}-x_0}{s}$$

$$SD = \frac{\sigma}{\sqrt{n}}, SE = \frac{s}{\sqrt{n}}$$

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

Large sample size is in fact increasing the probability of accepting the null hypothesis for e.g. $\alpha = 5\%$.

6 OLS, GLS, Feasible GLS

-OLS

$$\hat{\beta} = (X'X)^{-1}X'Y$$
$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon$$
$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = \sum_i e_i^2 = \nu S^2$$
$$E(R) = \mu$$
$$Var(\hat{\mu}) = \frac{\sigma^2}{T}$$
$$Var(\hat{\sigma}) = \frac{\sigma^2}{2T}$$
$$Var(\hat{\sigma^2}) = \frac{2\sigma^4}{T}$$

-Variance of Skewness:

$$Var(\hat{S_k}) \approx \frac{6}{T}$$

-Variance of Kurtosis:

$$Var(\hat{K}) \approx \frac{24}{T}$$

7 MLE

Likelihood is joint density.
MLE: F.O.C⇒

$$\hat{\beta}_{MLE} = (X'X)^X Y$$
$$\hat{\sigma}_{MLE} = \sqrt{\frac{1}{T}\epsilon'\epsilon}$$

MLE is invariant to transformation:

$$\widehat{f(\theta)} = f(\hat{\theta})$$

$$Var(\hat{\beta}_{MLE}) = (X'X)^{-1}\sigma^2$$

8 Regression Analysis

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

$$SS_{tot} = \sum_i = (y_i - \bar{y})^2$$

$$SS_{res} = \sum_i = (y_i - f_i)^2 = \sum_i e_i^2$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

9 Bayesian

$$p(\beta|\sigma) \sim N(\bar{\beta}, \sigma^2 A^{-1})$$

$$p(\sigma) \sim IG(\gamma_0, \gamma_0 S_0^2)$$

$$BF_{0/1} = \frac{P(M_0|y)}{P(M_1|y)} = \frac{\int P(y|M_0, \theta_0)P(M_0, \theta_0)d\theta_0}{\int P(y|M_1, \theta_1)P(M_1, \theta_1)d\theta_1}$$

9.1 Ratio of marginal likelihood:

Savage ratio test:

$$BF_{0/1} = \frac{Pesterior}{Prior}, \quad BF_{1/2} = \frac{BF_{1/3}}{BF_{2/3}}$$

10 MISC

Leibniz Rule:

$$P(Y(y_0)) = P_x(g^{-1}(y_0))|\frac{d}{dy}g^{-1}(y_0)|$$

Chebyshev:

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

Central Limit Theorem:

$$\lim_{n\to\infty} P(\frac{\bar{x_n}-\mu}{\sigma/\sqrt{n}} < k) = \Phi???$$

$$MSE[(\tilde{\theta}-\theta)^2] = (E(\tilde{\theta}-\theta))^2 - Var(\tilde{\theta})$$

CauchySchwarz inequality:

$$| < u, v > | \leq ||u|| \cdot ||v||$$

Limits:

$$(1+\frac{x}{n})^n = e^x (n \to \infty)$$

11 Logistic Regression

$$\begin{aligned} P(success) &= \pi \, (probability \, between \, 0 \, and \, 1) && (1) \\ P(failure) &= 1-\pi && (2) \\ Odds &= \frac{\#successes}{\#failures} = \frac{\pi}{1-\pi} && (3) \end{aligned}$$

11.1 logit function(Taking logit):

Dependent Variable, "The logit"(The log of Odds): $ln(\frac{\pi}{1-\pi})$

$$ln(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...$$

So we get back to linear regression, and:

$$\pi = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots)}}$$

11.2 Interpreting coefficients:

$$H_0 : Slope = 0$$

$$H_1 : Slope \neq 0$$

$$odds\, ratio = e^{\hat{\beta}_1}$$

Rescale:

$$e^{5\hat{\beta}_1}$$