Binomial Distribution

$$p(x) = \binom{n}{x} p^{x} (1-p)^{n-x} \qquad x \ge 0$$

$\mu = np$, $\sigma^2 = np(1-p)$ $\qquad Bi(n,p) \approx N(np, np(1-p))$ for large enough n

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

Uniform Distribution

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0 & \text{o.w.} \end{cases} \qquad \mu = \frac{1}{2}(a+b)$$

$$\sigma^2 = \frac{1}{12}(b-a)^2$$

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$$

Sampling Distributions (normal population)

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

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

- for paired sample, assume $D_i \sim N(\mu_D, \sigma_D^2)$ where $D_i = X_i - Y_i$

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \approx N(0,1) \text{ for large } n, m$$

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_P \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \qquad S_p = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}$$

$$\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim t_r \qquad r = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}$$

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

$$\frac{S_Y^2/\sigma_Y^2}{S_X^2/\sigma_X^2} = \frac{\left[\frac{(m-1)S_Y^2}{\sigma_Y^2}\right]/(m-1)}{\left[\frac{(n-1)S_X^2}{\sigma_X^2}\right]/(n-1)} \sim F_{m-1,n-1}$$

$\hat{p} = \bar{X}$, for large n; $\quad \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0,1)$

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \approx N(0,1) \qquad \bar{X} - X^* \sim N\left(0, 1 + \frac{1}{n}\right)$$

Sample sizes

$$\bar{x} \pm c\frac{\sigma}{\sqrt{n}} = \bar{x} \pm \epsilon \qquad c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$$\epsilon = c\frac{\sigma}{\sqrt{n}} \qquad n = \left(\frac{c\sigma}{\epsilon}\right)^2$$

$$\hat{p} \pm c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$$\epsilon = c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad n = \frac{c^2 \hat{p}(1-\hat{p})}{\epsilon^2}$$

$\hat{p}(1-\hat{p}) \to \frac{1}{4}$ as $\hat{p} \to 0.5$ $\qquad \hat{p}(1-\hat{p}) \le \frac{1}{4}$

$n = c^2/(4\epsilon^2)$ is a conservative choice

Regression

A regression model is linear if the predictor function is an additive function of x functions with the coefficients multiplied out the front.

$E(Y|x) = \alpha + \beta x = \alpha_0 + \beta(x - \bar{x})$ where $\alpha_0 = \alpha + \beta\bar{x}$

OLS estimators:

$$\hat{\alpha}_0 = \bar{Y}, \quad \hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \qquad \hat{\alpha} = \hat{\alpha}_0 - \hat{\beta}\bar{x} = \bar{Y} - \hat{\beta}\bar{x}$$

$$\hat{\mu}(x) = \hat{\alpha} + \hat{\beta}x = \hat{\alpha}_0 + \hat{\beta}(x - \bar{x}) = \bar{Y} + \hat{\beta}(x - \bar{x})$$

$$E(\hat{\alpha}_0) = E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^n E(Y_i) = \frac{1}{n}\sum_{i=1}^n [\alpha_0 + \beta(x_i - \bar{x})] = \alpha_0$$

$$E(\hat{\beta}) = \sum_{i=1}^n \frac{(x_i - \bar{x})}{K} E(Y_i) = \frac{1}{K}\sum_{i=1}^n (x_i - \bar{x})(\alpha_0 + (x_i - \bar{x})\beta)$$

$$= \frac{1}{K}\sum_{i=1}^n (x_i - \bar{x})\alpha_0 + \frac{K}{K}\beta = \beta$$

$\Rightarrow E(\hat{\alpha}) = \alpha$ and $E(\hat{\mu}(x)) = \mu(x)$

$$var(\hat{\alpha}_0) = var(\bar{Y}) = \frac{\sigma^2}{n}$$

$$var(\hat{\beta}) = \frac{\sigma^2}{K} \qquad var(\hat{\alpha}) = \left(\frac{1}{n} + \frac{\bar{x}^2}{K}\right)\sigma^2$$

$$cov(\hat{\alpha}_0, \hat{\beta}) = 0$$

$$var(\hat{\mu}(x)) = \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{K}\right)\sigma^2$$

$$\hat{\sigma}^2 = \frac{1}{n-2}D^2 \quad \text{where } D^2 = \sum_{i=1}^n (Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2$$

$\hat{Y}_i = \hat{\alpha}_0 + \hat{\beta}(x_i - \bar{x})$, $\quad R_i = Y_i - \hat{Y}_i$

$$D^2 = \sum_{i=1}^n R_i^2$$

substitute $\hat{\sigma}^2$ into older formulae of estimators in order to calculate standard errors

OLS estimates = MLE when we assume a normal population distribution

Regression (cont.)

$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{K}\right)$ $\qquad \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$

$\frac{\hat{\beta} - \beta}{\hat{\sigma}/\sqrt{K}} \sim t_{n-2}$ $\qquad \frac{\hat{\mu}(x) - \mu(x)}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{K}}} \sim t_{n-2}$

$\hat{\alpha}_0$ and $\hat{\beta}$ are independent because they are bivariate normal rvs with zero covariance

$$\hat{\mu}(x^*) \pm \underbrace{c}_{t_{n-2}}\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{K}}$$

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum x_i y_i - n\bar{x}\bar{y}\right)$$

$R = R_{xy} = \frac{S_{xy}}{S_x S_y}$ $\qquad R^2_{xy} = R^2 = R^2_{yx}$

$g(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \operatorname{artanh}(r)$ $\qquad g(R) \approx N\left(g(\rho), \frac{1}{n-3}\right)$

Distribution-free methods

Sign test: assume x is continuous $\qquad Y = $ # values greater than m_0

Under H_0, $Y \sim Bi(n, 0.5)$

Wilcoxon one-sample test: also assume distribution is symmetrical

W = sum of signed ranks of values vs m_0

for large n, $\frac{W - 0}{\sqrt{n(n+1)(2n+1)/6}} \approx N(0,1)$

for two samples, order the combined sample and let W be the sum of the ranks of Y_i

if $m_x > m_y$ we expect W to be small so we test for $W \leq c$ for this example.

$W \approx N\left(\frac{n_y(n_x+n_y+1)}{2}, \frac{n_x n_y(n_x+n_y+1)}{12}\right)$ for large n_x and n_y

Goodness-of-fit test / chi squared test:

k classes/possible outcomes

p_i = probability of the i^{th} class $\left(\sum_{i=1}^{k} p_i = 1\right)$

$$Q_{k-1} = \sum_{i=1}^{k}\frac{(Y_i - np_i)^2}{np_i} = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i} \approx \chi^2_{k-1}$$

reject H_0 if $Q_{k-1} > c$ where c is the $1-\alpha$ quantile from $\chi^2_{k-\text{est. param}-1}$

note Q measures badness of fit

Tests of independence (contingency tables)

$p_{i\cdot} = \sum_{j=1}^{c} p_{ij} = Pr(A_i)$ $\qquad p_{\cdot j} = \sum_{i=1}^{r} p_{ij} = Pr(B_j)$

$H_0: p_{ij} = p_{i\cdot}\, p_{\cdot j}$ $\qquad \hat{p}_{i\cdot} = \frac{y_{i\cdot}}{n}$ $\qquad \hat{p}_{\cdot j} = \frac{y_{\cdot j}}{n}$

$$Q = \sum_i \sum_j \frac{(Y_{ij} - np_{ij})^2}{np_{ij}} \approx \chi^2_{(r-1)(c-1)}$$

under H_0,

$$\hat{p}_{ij} = \hat{p}_{i\cdot}\,\hat{p}_{\cdot j} = \frac{Y_{i\cdot}\,Y_{\cdot j}}{n^2}$$

ANOVA

$\bar{X}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ $\qquad \bar{X}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} = \frac{1}{n}\sum_{i=1}^{k} n_i \bar{X}_{i\cdot}$

$$SS(TO) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{\cdot\cdot})^2 = SS(T) + SS(E)$$

between

$$SS(T) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 = \sum_{i=1}^{k} n_i(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$$

within

$$SS(E) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2 = \sum_{i=1}^{k}(n_i - 1)S_i^2$$

$\frac{SS(E)}{\sigma^2} \sim \chi^2_{n-k}$, $\hat{\sigma}^2 = \frac{SS(E)}{n-k} = MS(E)$ *true whether H_0 is true or not

$\frac{SS(TO)}{\sigma^2} \sim \chi^2_{n-1}$, $\hat{\sigma}^2 = \frac{SS(TO)}{(n-1)}$

Under H_0, $\bar{X}_{i\cdot} \sim N\left(\mu, \frac{\sigma^2}{n_i}\right)$ $\qquad MS(T) = \frac{SS(T)}{k-1}$

$\frac{SS(T)}{\sigma^2} \sim \chi^2_{k-1}$ and is independent of $SS(E)$

Under H_0, $\frac{E(SS(T))}{k-1} = \sigma^2$, otherwise $\frac{E(SS(T))}{k-1} > \sigma^2$

$$\frac{E(SS(E))}{n-k} = \sigma^2$$

$$F = \frac{SS(T)/(k-1)}{SS(E)/(n-k)} = \frac{MS(T)}{MS(E)} \sim F_{k-1,\, n-k}$$

reject H_0 if $F > c$

$F_A = \frac{SS(A)/(a-1)}{SS(E)/((a-1)(b-1))} \sim F_{a-1,\,(a-1)(b-1)}$ $\qquad \mu_{ij} = \mu + \alpha_i + \beta_j$

$F_{AB} = \frac{SS(AB)/[(a-1)(b-1)]}{SS(E)/[ab(c-1)]} \sim F_{(a-1)(b-1),\, ab(c-1)}$ $\qquad H_{0AB}: \gamma_{ij} = ($

$MS(E) = \frac{SS(E)}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \hat{\sigma}^2$ $\qquad$ manipulat[e] λ to get a known distrib[ution]

$MS(R) = \frac{SS(R)}{1} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$ $\qquad \lambda = \frac{L_0}{L_1}$ $\qquad \lambda < k$

$F = \frac{MS(R)}{\ldots}$

Binomial Distribution

$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x \geq 0$

$\mu = np, \ \sigma^2 = np(1-p)$ $\quad Bi(n,p) \approx N(np, np(1-p))$ for n large enough

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

Exponential Distribution

$f(x) = \lambda e^{-\lambda x} \quad F(x) = 1 - e^{-\lambda x}, \ \lambda > 0$

$\mu = \frac{1}{\lambda}, \ \sigma^2 = \frac{1}{\lambda^2}$

if $X_i \sim Exp(\lambda)$

$\sum X_i = n\bar{x} \sim Gamma(n,\lambda)$

$\lambda n\bar{x} \sim Gamma(n,1)$

$2\lambda n\bar{x} \sim \chi^2_{2n}$

Normal Distribution

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \mu = \mu \quad \sigma^2 = \sigma^2$

$\bar{x} \xrightarrow{d} N(\mu, \frac{\sigma^2}{n}), \quad \hat{p} = \frac{x}{n} \xrightarrow{d} N(p, \frac{p(1-p)}{n})$

$x \sim Bi(n,p)$

Uniform Distribution

$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0 & \text{otherwise} \end{cases} \quad \mu = \frac{1}{2}(a+b)$

$\sigma^2 = \frac{1}{12}(b-a)^2$

$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$

Order statistics and quantiles

$Pr(X_{(4)} \leq 0.5)$ = Pr(at least 4 X_is less than 0.5)

= Pr(4 X_is less than 0.5)

+ Pr(5 X_is less than 0.5)

$= \binom{5}{4} 0.25^4 0.75 + 0.25^5$

$g_k(x) = k\binom{n}{k} F(x)^{k-1} (1-F(x))^{n-k} f(x)$

$g_1(x) = n(1-F(x))^{n-1} f(x)$

$g_n(x) = n F(x)^{n-1} f(x)$

$Pr(X_{(1)} > x) = (1-F(x))^n, \quad Pr(X_{(n)} \leq x) = F(x)^n$

$p = F(\pi_p) = Pr(X \leq \pi_p), \quad \pi_p = F^{-1}(p)$

Type 1: $\hat{\pi}_p = x_{(\lceil np \rceil)} = x_{(k)}$ if $\frac{k-1}{n} < p \leq \frac{k}{n}$

Type 6: $\hat{\pi}_p = x_{(k)}$, where $p = \frac{k}{n+1}$

Type 7: $\hat{\pi}_p = x_{(k)}$, where $p = \frac{k-1}{n-1}$

$g_k(w) = k\binom{n}{k} w^{k-1}(1-w)^{n-k}, \quad W_k = F(X_{(k)})$

$F(X_k) \sim Beta(k, n-k+1), \quad E(W_k) = \frac{k}{n+1}$

$mode(W_k) = \frac{k-1}{n-1}$

$\hat{\pi}_p \approx N\left(\pi_p, \frac{p(1-p)}{nf(\pi_p)^2}\right) \quad \hat{M} \approx N\left(m, \frac{1}{4nf(m)^2}\right)$

for large n, where $f(x)$ is population pdf

W = number of $X_i < m$, $W \sim Bi(5, 0.5)$

$Pr(X_{(k)} \leq m) = \ldots$

Bayesian methods

$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{Pr(A)} \quad Pr(A) = Pr(A \cap B) + Pr(A \cap B')$

$= Pr(A|B)Pr(B) + Pr(A|B')Pr(B')$

$Pr(B_i|A) \propto Pr(A|B_i)Pr(B_i)$

$\underbrace{Pr(\theta|X)}_{\text{posterior}} = \frac{\overbrace{Pr(X=x|\theta=a)}^{L(\theta)} \overbrace{Pr(\theta=a)}^{\text{prior}}}{Pr(X=x)}$

$Pr(X=x) = Pr(X=x|\theta=\theta_1)Pr(\theta=\theta_1)$

$+ Pr(X=x|\theta=\theta_2)Pr(\theta=\theta_2) \ldots$

$Pr(\theta=a|X=x) \propto Pr(X=x|\theta=a)Pr(\theta=a)$

Example

$X \sim Bi(n,\theta)$ $\quad$ uniform prior

$\theta \in [0,1]$ $\quad f(\theta) = 1, \ 0 \leq \theta \leq 1$

$f(\theta|X=x) \propto Pr(X=x|\theta) f(\theta)$

$\propto \theta^x (1-\theta)^{n-x}$

$1 = \int_0^1 f(\theta|X=x) = \int_0^1 A\theta^x(1-\theta)^{n-x} d\theta$

$A = \frac{1}{\int_0^1 \theta^x (1-\theta)^{n-x} d\theta}$ ← normalising constant

$= \frac{x!(n-x)!}{(n+1)!}$ so $\sim \beta(x+1, n-x+1)$

Beta prior + binomial likelihood ⇒ beta posterior

Beta prior is a conjugate prior for the binomial distribution

Uniform prior ⇒ posterior mode = MLE

$sd(\theta|X=x)$ is a measure of uncertainty (std error)

Example:

Random sample: $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ with σ^2 known

Summarise the data by $Y = \bar{X} \sim N(\theta, \frac{\sigma^2}{n})$

Prior: $\theta \sim N(\mu_0, \sigma_0^2)$

Posterior: $f(\theta|y) f(\theta)$

$= \frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2/n}(y-\theta)^2}$

$\cdot \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(\theta-\mu_0)^2}$

$\exp\left[-\frac{(y-\theta)^2}{2\sigma^2/n} - \frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right]$ expand out and combine quadratics

$f(\theta|y) \propto \exp\left[-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}\right]$ complete the square

$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{y}{\sigma^2/n}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}} \quad \frac{1}{\sigma_1^2} = \underbrace{\frac{1}{\sigma_0^2}}_{\text{prior precision}} + \underbrace{\frac{1}{\sigma^2/n}}_{\text{data precision}}$

improper prior: cannot integrate to 1

$\sigma_1^2 = \sigma^2/n \quad \mu_1 = y = \bar{x}$ (posterior dominated by data)

Asymptotics & optimality

MLE is asymptotically:

- unbiased
- efficient (has the optimal variance)
- normally distributed

$$U(\theta) = \frac{\partial l}{\partial \theta} \qquad V(\theta) = -\frac{\partial U}{\partial \theta} = -\frac{\partial^2 l}{\partial \theta^2}$$

$E(U(\theta)) = 0, \quad E(V(\theta)) = I(\theta)$

$var(U(\theta)) = I(\theta)$

MLE $\hat{\theta} \approx N\left(\theta, \frac{1}{I(\theta)}\right)$ as $n \to \infty$

$se(\hat{\theta}) = \frac{1}{\sqrt{I(\hat{\theta})}}$ or $\frac{1}{\sqrt{V(\hat{\theta})}}$ if we don't know $I(\theta)$

Example:

$x_1, \dots, x_n$ random sample from

$$f(x|\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty, \quad 0 < \theta < \infty$$

MLE is $\bar{X}$ (log of pdf for one observation)

$\ln f(x|\theta) = -\ln\theta - x/\theta$

$$U_i(\theta) = \frac{\partial}{\partial\theta} \ln f(x|\theta) = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

$$V_i(\theta) = -\frac{\partial^2}{\partial\theta^2} \ln f(x|\theta) = -\frac{1}{\theta^2} + \frac{2x}{\theta^3}$$

$$I_i(\theta) = E(V_i(\theta)) = E\left(-\frac{1}{\theta^2} + \frac{2X}{\theta^3}\right) = -\frac{1}{\theta^2} + \frac{2\theta}{\theta^3} = \frac{1}{\theta^2}$$

$I(\theta) = \frac{n}{\theta^2}, \quad \hat{\theta} \simeq N\left(\theta, \frac{\theta^2}{n}\right)$

C.R.L.B: $var(\hat{T}) \geq \frac{1}{I(\theta)}$

$$eff(T) = \frac{1/I(\theta)}{var(T)} = \frac{1}{I(\theta)\, var(T)}$$

$0 \leq eff(T) \leq 1$

$Y = g(x_1, \dots, x_n)$ is sufficient for θ if and only if

$$f(x_1, \dots, x_n | \theta) = \phi\{g(x_1, \dots, x_n) | \theta\}\, h(x_1, \dots, x_n)$$

ϕ depends on $x_1, \dots, x_n$ only through $g(x_1, \dots, x_n)$ and h doesn't depend on θ

$\sum_{i=1}^{n} K(X_i)$ is sufficient for θ

if $f(x|\theta) = \exp\{K(x)p(\theta) + S(x) + q(\theta)\}$

Uniformly most powerful tests:

$H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$

LRT is the most powerful test for a given significance level

$H_1 : \theta \in A_1$

If the form of the LRT differs for different values of θ_1, then any given one will only be the best for particular values of θ_{10}

(i.e. uniformly most powerful for $\theta_1 \in A_1$)