

Contents

Introduction

Probability

Odds, Log-Odds

Let $\pi = P(A)$. We define the "Odds Ratio" $\theta = \frac{P(A)}{1-P(A)} = \frac{\pi}{1-\pi}$. If $\theta = 3$ we say that the patient is three times more likely to have the disease. We also symmetrize this into "evidence" via the log odds definition:

$$L = \text{Logit}(\theta) = \log(\theta) = \log\left(\frac{\pi}{1-\pi}\right), \quad L \in (-\infty, \infty).$$

The units are either Bels (base 10) or nepers (base e). We can return to probabilities by taking the inverse

$$\pi = P(A) = \frac{e^L}{1+e^L} \quad \text{and} \quad 1-\pi = P(A^c) = \frac{1}{1+e^L}.$$

Moments

Calculations involving population/sample means and variances, and inner/outer products:

Population – [Sample size $n=1$]. **NOTE:** $\theta = (E(X), \text{Var}(X)) = (\mu'_1, \mu_2) = (" \mu ", " \sigma^2 ") \neq (\bar{X}_n, S_n^2)!$

$$\text{Var}(X) = \text{Cov}(X, X) = E(X - \mu)^2 = E(X^2) - \mu^2 \doteq \mu_2 \neq \sum_{k=1}^n (x_k - \mu)^2 \neq \sum (x - \bar{x})^2 \quad X \in R^{d=1} \quad (n=1)$$

$$\text{unless } X \text{ is a discrete r.v. and } f_X(x) \equiv \frac{1}{n} \forall x \in \text{Supt}(f).$$

$$\text{Var}(X) = \text{Cov}(X, X) = E(X - \mu)(X - \mu)^T = EXX^T - \mu_X \mu_X^T \doteq \Sigma_X \quad X \in R^d \quad (n=1)$$

$$\text{Cov}(X, Y) = E(X - \mu_X)(Y - \mu_Y)^T = EXY^T - \mu_X \mu_Y^T \doteq \Sigma_{XY} \quad (dxk), \quad X \in R^d \quad Y \in R^k \neq \sum_{n=1}^N \text{etc.}$$

Sample - [size $n \geq 1$] $m_k \equiv \hat{\mu}'_k = \frac{1}{n} \sum_{i=1}^n X_i^k$; $m_1 \equiv \bar{X} = \hat{\mu}'_1$; $m_2 - m_1^2 = S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

$$\sum_{k=1}^n (x_k - \mu)^2 = \sum (x - \bar{x})^2 + n(\bar{x} - \mu)^2$$

$$\sum_{k=1}^n (X_k - \bar{X})^2 = \sum (X^2 - 2X\bar{X} + \bar{X}^2) = \sum X^2 - n\bar{X}^2 = nS_1^2 = (n-1)S^2 \text{ (} S^2 \text{ unbiased)}$$

NOTE: $\sum x^2 = \sum (x - \bar{x})^2 + n\bar{x}^2$ $S^2 = \left(\frac{1}{n-1} \right) \sum X^2 - \left(\frac{n}{n-1} \right) \bar{X}^2 = \frac{n}{n-1} S_1^2$ $S_1^2 = \frac{n-1}{n} S^2$

$$\sum_{k=1}^n (X_k - \bar{X})(X_k - \bar{X})^T = \sum XX^T - n\bar{X}\bar{X}^T = n\hat{\Sigma}_x \text{ (} dx \times d \text{)} \quad X_k \in R^d, \quad k=1, n$$

$$n\hat{\Sigma}_x = n \begin{bmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1d} \\ s_{21} & s_{22} & s_{23} & \cdots & s_{2d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{d1} & s_{d2} & s_{d3} & \cdots & s_{dd} \end{bmatrix} \quad n\hat{\Sigma}_{ij} = nS_{ij} = \sum_k x_i x_j - n\bar{x}_i \bar{x}_j \quad i, j=1, d \quad k=1, n$$

For $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $n\hat{\Sigma}_x = \begin{bmatrix} \sum_k x_1^2 - n\bar{x}_1^2 & \sum x_1 x_2 - n\bar{x}_1 \bar{x}_2 \\ \sum x_2 x_1 - n\bar{x}_2 \bar{x}_1 & \sum x_2^2 - n\bar{x}_2^2 \end{bmatrix} = \begin{bmatrix} \sum_k x_1^2 - n\bar{x}_1^2 & \sum x_1 x_2 - n\bar{x}_1 \bar{x}_2 \\ \sum x_1 x_2 - n\bar{x}_1 \bar{x}_2 & \sum x_2^2 - n\bar{x}_2^2 \end{bmatrix} \quad k=1, n$

$$\sum_{k=1}^n (X - \bar{X})(Y - \bar{Y})^T = \sum XY^T - n\bar{X}\bar{Y}^T = n\hat{\Sigma}_{XY}, \quad X \in R^d, \quad Y \in R^k, \quad \hat{\Sigma}_{XY} \text{ (} dx \times dk \text{)}$$

Calculations involving population/sample higher-order moments (3rd and 4th), and tensor notation:

Population - [n=1]. NOTE:

Skew[ness]

Pearsonian notation: Skew “Sk” = $\frac{d}{\mu_2^{1/2}} = \frac{d}{\sigma}$ where $d = x_{me} - x_{mo}$, the distance from mean to mode. He also

used $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$. Later, “Modern” notation for skewness became $\gamma_1 \triangleq \tilde{\beta}_1 = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{\mu_3}{\sigma^3}$.

Recall that $\mu_3 = E(X - \mu)(X - \mu)(X - \mu) = \int (X - \mu)^3 f(x | \theta) dx$. For vector X, tensor notation is required to generalize.

Kurtosis

$\beta_2 = \frac{\mu_4}{\mu_2^2}$. Since $\beta_2 = 3$ for the normal distribution, Pearson also used an “excess kurtosis” $\tilde{\beta}_2 = \beta_2 - 3$. This

now has a modern notation of $\gamma_2 = \beta_2 - 3$. Note that in terms of the cumulants, $\gamma_2 = \frac{\kappa_4}{\kappa_2^2}$.

Sample - [size $n \geq 1$]

The skewness and kurtosis are normally estimated using method of moments. Standard errors for these statistics for both the normal and the Pearson system of distributions are provided in Pearson, Filon, etc. [1895, 1900, 1903, 1907]. Sample versions of d require mode estimation of determination of a Pearson curve since many mode estimators are highly sensitive to sampling variation.

More later

Non-independence of higher moments

$\hat{\beta}_1$ and $\hat{\beta}_2$ are necessarily not independent of themselves and the usual estimators of μ'_1 and μ_2 .

Covariance \Leftrightarrow Correlation

These identities require creating a $d \times d$ diagonal matrix with the variances as elements. These can be generated by first creating a variance vector, and by using the unit matrices $E_i : E = 0_{d \times d}, \{E_{ii}\} = 1$ and $e_i : e = 0_d, \{e_i\} = 1$. Note that the total number of covariances (including the diagonal variances) equals

$$d + \frac{d(d-1)}{2}.$$

Population

$$V = \sum_{i=1}^d E_i v e_i^T \text{ where } v = \text{diag}(\Sigma) = (\sigma_{11}, \dots, \sigma_{dd}) \text{ or } \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_d^2 \end{pmatrix}$$

$$P = V^{-\frac{1}{2}} \Sigma V^{-\frac{T}{2}} = V^{-\frac{1}{2}} \Sigma V^{-\frac{1}{2}} \quad \rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

$$\Sigma = V^{\frac{1}{2}} P V^{\frac{1}{2}}$$

$$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j$$

Note: In Matlab and R, $V^{\frac{1}{2}} = \text{chol}(V)$

Sample

$$S = \sum_{i=1}^d E_i s e_i^T \text{ where } s = \text{diag}(\hat{\Sigma})$$

$$R = S^{-\frac{1}{2}} \hat{\Sigma} S^{-\frac{1}{2}}$$

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}} \sqrt{s_{jj}}} = \frac{s_{ij}}{s_i s_j}$$

$$\hat{\Sigma} = S^{\frac{1}{2}} R S^{\frac{1}{2}}$$

$$s_{ij} = r_{ij} s_i s_j, \text{ or } \hat{\sigma}_{ij} = r_{ij} \hat{\sigma}_i \hat{\sigma}_j$$

Distributional Foundations

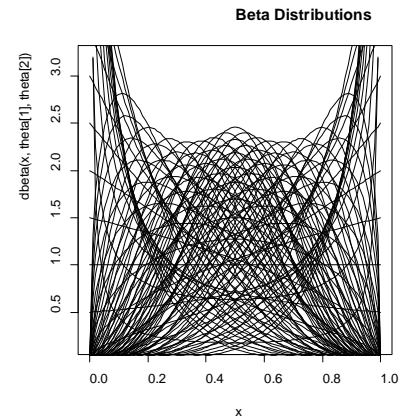
"Brand Name" Distributions

See for example [my Rice website](#).

- Beta(α, β):

Distributional Relationships and Identities

\equations\distributions\handy.relationships.wmf



- $F_X(k) = \sum_{x=0}^k \text{binom}(x | n, p) = \int_0^{1-p} \text{Beta}(y | n-k, k+1) dy = F_Y(1-p)$
- If $z \sim \phi(z) = N(0, 1)$, $z^2 \sim \chi_1^2$, or $\sum z^2 = z^T z \sim \chi_n^2$. For $\sigma z = \varepsilon \sim N(0, \sigma^2)$, $\frac{1}{\sigma^2} \varepsilon^T \varepsilon \sim \chi_n^2$.
- $\text{gamma}(n, 2) \stackrel{d}{=} \chi_{2n}^2$, or $\text{gamma}(n/2, 2) \stackrel{d}{=} \chi_n^2$; so if $X \sim \text{gamma}(n, \beta)$, $Y = (2/\beta)X \sim \text{gamma}(n, 2) \stackrel{d}{=} \chi_{2n}^2$
- $Y \sim \text{gamma}(\frac{\nu}{2}, \frac{\nu}{2})$ and $X | y \sim N_d(\mu, y^{-1}\Sigma)$, then $X \sim t_d(\mu, \Sigma, \nu)$
- $X \sim \text{gamma}(\alpha, \beta)$, $Y \sim \text{Pois}(x/\beta)$, then $P(X < x) = P(Y \geq \alpha)$.
- $X \sim \text{Pois}(\theta)$, $E(S^2 \cdot (\bar{X})^\delta) = E((\bar{X})^{\delta+1})$, $\delta \geq 0$
- $X \sim \text{expo}(\beta)$ then $Y = X^{1/\gamma} \sim \text{Weibull}(\gamma, \beta)$
- $X \sim \text{expo}(\beta)$ then $Y = \sqrt{\frac{2X}{\beta}} \sim \text{Rayleigh}$

- $X \sim \text{expo}(\beta)$ and $\beta = 1$ then $Y = \alpha - \gamma \ln X \sim \text{Gumbel}(\alpha, \gamma), \alpha \in \mathbb{R}, \gamma > 0$
- $X \sim \text{expo}(\lambda)$ and $Y \sim \text{Erlang}(n, \lambda)$ then $X / Y \sim \text{Pareto}(1, n)$
- $X \sim \text{gamma}(1, \beta) \stackrel{d}{=} \text{expo}(\beta)$
- $X \sim \text{gamma}(\alpha, \beta)$ then $1 / X \sim \text{IG}(\alpha, \beta)$
- $X \sim \text{gamma}(1.5, \beta)$ then $Y = \sqrt{X / \beta} \sim \text{Maxwell}$
- $X \sim \text{gamma}(\alpha, \theta), Y \sim \text{gamma}(\beta, \theta)$ then $\frac{X}{X+Y} \sim \text{beta}(\alpha, \beta)$
- $X \sim \chi^2(\alpha), Y \sim \chi^2(\beta)$ then $\frac{X}{X+Y} \sim \text{beta}(\frac{\alpha}{2}, \frac{\beta}{2})$.
- $X \sim \text{unif}(0, 1), \alpha > 0$ then $X^{1/\alpha} \sim \text{beta}(\alpha, 1)$
- $X \sim \text{unif}(0, 1) \stackrel{d}{=} \text{beta}(1, 1)$
- $X \sim \text{beta}(1, 1)$ then $-\ln X \sim \text{expo}(1)$ and if $X \sim \text{beta}(\alpha, 1)$ then $-\ln X \sim \text{expo}(\alpha)$
- $X \sim \text{beta}(\alpha, \beta)$ then $1 - X \sim \text{beta}(\beta, \alpha)$
-

Sampling and Certain Statistics

Quantile function

The Quantile function is the inverse of the d.f.:

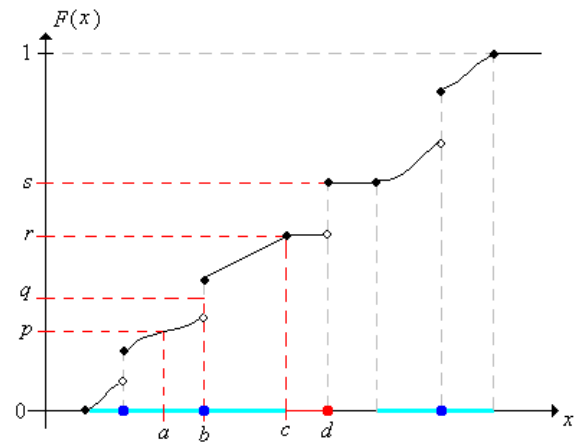
$$\xi_p = F_X^{-1}(p) = \inf\{x : F(x) \geq p\}$$

$$\xi_p = a$$

$$\xi_q = b$$

$$\xi_r = c$$

$$\xi_s = d$$



Some Sampling Facts

- \bar{X}, S^2 ind. $\Leftrightarrow X_i \sim \text{Normal}$; moreover, $\{X_i\}_{i=1}^n$ must be a SRS.
- $E(S^2) = \sigma^2$

pf:

$$\begin{aligned} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) &= E\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) \\ &= n(E(X_1^2) - E(\bar{X}^2)) \\ &= n\left(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2\right) = (n-1)\sigma^2 \end{aligned}$$

- $E(S) \leq \sigma$

pf:

$$\text{Var}(Y) = E(Y^2) - E(Y)^2 \geq 0$$

$$E(Y^2) \geq E(Y)^2. \text{ Let } Y = S$$

$$\sigma^2 = E(S^2) \geq E(S)^2, \text{ or } E(S) \leq \sigma$$

- In general, $\text{Var } S^2 = \frac{1}{n} \left(\mu_4 - \mu_2^2 \frac{(n-3)}{(n-1)} \right) \approx \frac{1}{n} (\mu_4 - \mu_2^2 + 4\mu_2\mu^2 - 4\mu\mu_3)$

$$\underline{X_i \sim N(\mu, \sigma^2):} \quad \mu_2^2 = \sigma^4, \mu_4 = 3\sigma^4$$

$$\text{Var } S^2 = \frac{2\sigma^4}{(n-1)}$$

$$\underline{X_i \sim \text{Unif}(a, b):} \quad \mu_2 = \frac{(b-a)^2}{12}, \mu_4 = \frac{9}{5}$$

$$\text{Var } S^2 = \frac{1296(n-1) - 5(n-3)(b-a)^4}{720n(n-1)}$$

$$\underline{X_i \sim t_\nu:} \quad \mu_2 = \frac{\nu}{\nu-2}, \mu_4 = \frac{3\nu^2}{(\nu-2)(\nu-4)}$$

$$\text{Var } S^2 = \frac{2\nu^2(2\nu-5)}{n(\nu-2)^2(\nu-4)}$$

$$\underline{X_i \sim \chi_\nu^2:} \quad \mu_2 = 2\nu, \mu_4 = \frac{3(\nu+4)}{\nu}$$

$$\text{Var } S^2 = \frac{3(n-1)(\nu+4) - 4(n-3)\nu^3}{n(n-1)\nu}$$

$$\underline{X_i \sim \text{expo}(\theta):} \quad \mu_2 = \theta^2, \mu_4 = 9\theta^4$$

$$\text{Var } S^2 = \frac{2(4n-3)\theta}{n(n-1)}$$

$$\underline{X_i \sim \text{Poisson}(\lambda):} \quad \mu_2 = \lambda, \mu_4 = \lambda(3\lambda+1)$$

$$\text{Var } S^2 = \frac{2\lambda^2}{n-1} + \frac{\lambda}{n} = \frac{\lambda}{n} \left(\frac{2n\lambda}{n-1} + 1 \right)$$

Order Statistics

1 Dahiya and Gurland, Biometrics v.25, No.1 (1969) pp.171-173

2 Cho, E. & M. J. Cho, 2008. Variance of sample variance. *Proceedings of the 2008 Joint Statistical Meetings, Section on Survey Research Methods*, American Statistical Association, Washington DC, 1291-1293.