

Binomial $X \sim B(n, p)$

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$$

$$E X = np, \quad \text{Var } X = np(1-p)$$

$$\text{mgf: } M_X(t) = (pe^t + 1 - p)^n$$

$$\text{Beta is conjugate prior, Fisher info } I(p) = \frac{1}{p(1-p)}$$

Poisson $X \sim P(\lambda)$

$$p(k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad k = 0, 1, \dots$$

$$E X = \lambda, \quad \text{Var } X = \lambda$$

$$\text{mgf: } M_X(t) = e^{\lambda(e^t - 1)}$$

$$\text{Gamma is conjugate prior, Fisher info } I(\lambda) = \frac{1}{\lambda}$$

Normal $X \sim N(\mu, \Sigma)$, Σ positive definite

$$f(x) = \frac{\exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}}{(2\pi)^{\frac{k}{2}} \sqrt{\det(\Sigma)}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\text{mgf: } M_X(t) = \exp(\mu' t + \frac{1}{2} t' \Sigma t)$$

$$\text{Normal is conj. prior, Fisher info } I(\mu, \sigma^2) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/2\sigma^4 \end{bmatrix}$$

Beta $X \sim \text{Beta}(\alpha, \beta)$

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad 0 \leq x \leq 1$$

$$E X = \frac{\alpha}{\alpha+\beta}, \quad \text{Var } X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

using the beta function:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$$

Gamma $X \sim \text{Gamma}(\alpha, \beta)$

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad x > 0$$

$$E X = \frac{\alpha}{\beta}, \quad \text{Var } X = \frac{\alpha}{\beta^2}$$

$$\text{mgf: } M_X(t) = (1 - \frac{t}{\beta})^{-\alpha}, t < \beta$$

$$X \sim \text{Gamma}(\alpha, \beta) \iff \beta X \sim \text{Gamma}(\alpha, 1)$$

X_i iid $\text{Gamma}(\alpha_i, \beta)$, then

$$\sum X_i \sim \text{Gamma}(\sum \alpha_i, \beta)$$

$$\text{Gamma function: } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}.$$

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$$

$$\Gamma(k) = (k-1)! \text{ for } k \text{ positive integer.}$$

Exponential Special case: $\text{Exp}(\lambda) \equiv \text{Gamma}(1, \lambda)$

Chi square Special case: $\chi_n^2 \equiv \text{Gamma}(\frac{n}{2}, \frac{1}{2})$

Let Z_i be iid $N(0, 1)$.

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2$$

Noncentral χ^2 . Let $Y \sim N(\mu, I)$ be an n vector. Then

$$\|Y\|^2 \sim \chi_n^2(\|\mu\|^2)$$

F

$$F(m, n) \equiv \frac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}}$$

Where numerator and denominator are independent χ^2 .

T

$$t(n) = \frac{N(0, 1)}{\sqrt{\frac{\chi_n^2}{n}}}$$

Where numerator and denominator are independent.

Moment generating functions determine distribution

$$M_X(t) \equiv E(e^{tX}), \quad M'_X(0) = E(X)$$

X_i independently distributed \iff

$$M_{\sum X_i}(t) = \prod M_{X_i}(t)$$

Characteristic function

$$\phi(t) = E(e^{it^T X}) = E(\cos(t^T X)) + i E(\sin(t^T X))$$

Order statistics for sorted sample $X_{(1)}, \dots, X_{(n)}$ has pdf:

$$n! \prod_{i=1}^n f(X_{(i)}) \quad I(X_{(1)} < \dots < X_{(n)})$$

Jensen's Inequality if $S \subset R^k$ convex and closed, g convex on S , $P[X \in S] = 1$, and $E X$ is finite, then $E X \in S$, $E g(X)$ exists, and

$$E g(X) \geq g(E X)$$

Holder's Inequality if $r, s > 1$ and $\frac{1}{r} + \frac{1}{s} = 1$ then

$$E |XY| \leq (E |X|^r)^{\frac{1}{r}} (E |Y|^s)^{\frac{1}{s}}$$

$T(X)$ Sufficient means the distribution of $X|T(X)$ does not depend on θ .

Factorization theorem: $T(x)$ is sufficient \iff

$$f_{\theta}(x) = h(x)g(\theta, T(x))$$

$L_x(\theta) = p_{\theta}(x) = p(x, \theta)$ likelihood is function of θ , density is function of x .

The likelihood ratio

$$\lambda_x(\theta) = \frac{L_x(\theta)}{L_x(\theta_0)}$$

is minimal sufficient. To show $T(x)$ is minimal sufficient show that it is sufficient and a function of the likelihood $\lambda_x(\theta)$.

Fisher information

$$I(\theta) = E_{\theta} \left[\frac{\partial}{\partial \theta} \log L_X(\theta) \right]^2 = E_{\theta} \left[-\frac{\partial^2}{\partial \theta^2} \log L_X(\theta) \right]$$

bias $\hat{v} \equiv E(\hat{v}) - v$

$$MSE(\hat{v}) \equiv E(\hat{v} - v)^2 = \text{Var}(\hat{v}) + (\text{bias } \hat{v})^2$$

Rao-Blackwell Let $S(X)$ be an unbiased point estimator for $g(\theta)$. Conditioning on a sufficient statistic $T(X)$ reduces variance.

$$\text{Var}_{\theta}(S(X)) \geq \text{Var}_{\theta}(E(S(X)|T(X)))$$

Also holds for more general convex loss function L :

$$R(\theta, S) \equiv E_{\theta} L(\theta, S(X)) \geq E_{\theta} L(\theta, E(S(X)|T(X)))$$

Completeness $T(X)$ is complete if $E g(T(X)) = 0$ implies $g = 0$ almost surely for all θ .

Cramer Rao Inequality Let $g : \Theta \rightarrow R$. Suppose there exists an unbiased estimator $U(X)$, $E U(X) = g(\theta)$. Then

$$\text{Var}_{\theta} U(X) \geq \left(\frac{\partial g(\theta)}{\partial \theta} \right)^T I(\theta)^{-1} \left(\frac{\partial g(\theta)}{\partial \theta} \right)$$

Basu's Theorem - If $T(X)$ complete sufficient statistic and $A(X)$ is ancillary then $A(X)$ and $T(X)$ are independent.

Lehmann - Scheffe Suppose $T(X)$ is complete sufficient. Then there exists unique unbiased estimator $E h(T(x))$ of $g(\theta) \in R$ with smallest variance (MVUE).

Exponential Families $T(x)$ is natural sufficient statistic and is complete sufficient if the k parameter exponential family is full rank.

$$p(x, \theta) = h(x) \exp\{\eta(\theta)^T T(x) - B(\theta)\}$$

Canonical form model indexed by η .

$$q(x, \eta) = h(x) \exp\{\eta^T T(x) - A(\eta)\}$$

$$\dot{A}(\eta) = E_{\eta}(T(X)) \quad \ddot{A}(\eta) = I(\eta) = \text{Var}_{\eta}(T(X))$$

Then moment generating function for $T(X)$ is

$$M_{T(X)}(t) = \exp\{A(t + \eta) - A(\eta)\}$$

Equivalent statements useful for GLM's such as $Y \sim N(X\beta, \sigma_0^2 I)$, where Z is $n \times p$:

1. $I(\beta) = \frac{1}{\sigma_0^2} X^T X$ positive definite 2. $\text{rank}(X) = p$ 3. model is identifiable. More generally another equivalent statement is $\text{Var}(T(X)) = \ddot{A}(\eta)$ is positive definite.

Decision Theory

Decision rule $\delta : \mathcal{X} \rightarrow \mathcal{A}$, where $\delta \in \mathcal{D}$, the space of possible decision rules and \mathcal{A} is the action space.

Loss function $l : \Theta \times \mathcal{A} \rightarrow \mathbb{R}^+$ Posterior mean minimizes square error loss; median minimizes absolute loss.

Risk function $R : \Theta \times \mathcal{D} \rightarrow \mathbb{R}^+$ expected loss for a particular value of θ

$$R(\theta, \delta) = E_{\theta} l(\theta, \delta(X)) = \int l(\theta, \delta(x)) \cdot p_{\theta}(x) dx$$

Bayes setup:

$$\pi(\theta|x) = \frac{p_{\theta}(x)\pi(\theta)}{m(x)}$$

Bayes decision rule If there exists $\delta_{\pi} \in \mathcal{D}$ w.r.t prior π such that

$$r(\pi, \delta_{\pi}) = \inf_{\delta \in \mathcal{D}} r(\pi, \delta)$$

To find Bayes rule minimize the posterior risk:

$$\delta_{\pi}(x) = \min_{a \in \mathcal{A}} r_{\pi}(a|x)$$

Bayes risk $r_{\pi} : \mathcal{D} \rightarrow \mathbb{R}^+$ expected loss for fixed prior π

$$r_{\pi}(\delta) = E_{\pi} R(\theta, \delta) = \int_{\Theta} R(\theta, \delta) \pi(d\theta) = \int_{\mathcal{X}} r_{\pi}(\delta(x)|x) m(x) dx$$

To find Bayes risk: 1) find the Bayes rule 2) compute the risk function 3) take the expectation of the risk wrt prior π .

Minimax decision rule δ^* minimizes the worst case scenario, satisfies

$$\sup_{\theta} R(\theta, \delta^*) = \inf_{\delta} \sup_{\theta} R(\theta, \delta)$$

To show δ^* is minimax, first check for constant risk $R(\theta, \delta^* = c$ for all θ , then find a prior π such that δ^* is the Bayes rule. This π is least favorable. More generally can find a sequence of priors (π_k) such that the Bayes risk $r_{\pi_k}(\delta_{\pi_k}) \rightarrow c$.

Multivariate Normal

Stein's formula: $X \sim N(\mu, \sigma)$

$$E(g(X)(X - \mu)) = \sigma^2 E(g'(X))$$

assuming these expectations are finite.

$X \sim N(\mu, \Sigma)$, A an $m \times n$ matrix, then

$$AX \sim N(A\mu, A\Sigma A^t)$$

For Σ full rank it's possible to transform between $Z \sim N(0, I)$ and X :

$$X = \Sigma^{1/2}Z + \mu \quad Z = \Sigma^{-1/2}(X - \mu)$$

In block matrix form:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

Assuming Σ_{11} is positive definite then the conditional distribution

$$X_2|X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

Conditional Distributions

Conditional pdf:

$$f_{X|Y}(x, y) \equiv \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Iterated expectation:

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

Conditional variance formula:

$$\text{Var}(Y) = \text{Var}(E(Y|X)) + E(\text{Var}(Y|X))$$

General Techniques

Singular Value Decomposition (SVD) Any matrix X can be written

$$X = UDV^T$$

with U, V orthogonal, and D diagonal.

Moore Penrose Psuedoinverse A^+ exists uniquely for every matrix A .

Projection matrix P are symmetric and idempotent. They have eigenvalues either 0 or 1.

$$P = P^T \quad P^2 = P$$

Covariance of linear transformations

$$\text{Cov}(Ay, Bx) = A\text{Cov}(y, x)B^T$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Integration by parts:

$$\int uv' = uv - \int u'v$$

Transformation theorem: Given some regularity conditions, $Y = g(X)$ has pdf

$$p_Y(y) = p_X(g^{-1}(y))|J_{g^{-1}(y)}|$$

Matrix / Vector differentiation

$$\frac{\partial A^T \beta}{\partial \beta} = A, \quad \frac{\partial \beta^T A \beta}{\partial \beta} = (A + A^t)\beta = 2A\beta \text{ for } A \text{ symmetric.}$$