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

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

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

mgf: $M_X(t) = (pe^t + 1 - p)^n$

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

Poisson $X \sim P(\lambda)$

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

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

mgf: $M_X(t) = e^{\lambda(e^t - 1)}$ Use recursive relation to compute $\mathrm{E}(X_i)$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$f(x) = \lambda e^{-\lambda x}, \quad x > 0 \qquad \mathrm{E}\, X = \frac{1}{\lambda}, \quad \mathrm{Var}\, X = \frac{1}{\lambda^2}$

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

Chi square Special case: $X \sim \chi^2_n \equiv \mathrm{Gamma}(\frac{n}{2}, \frac{1}{2})$

$f(x) \propto x^{\frac{n}{2}-1} e^{\frac{-x}{2}}, \quad x > 0 \qquad \mathrm{E}\, X = n, \quad \mathrm{Var}\, X = 2n$

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

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

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

$$\|Y\|^2 \sim \chi^2_n(\|\mu\|^2)$$

F

$$F(m,n) \equiv \frac{\frac{\chi^2_m}{m}}{\frac{\chi^2_n}{n}}$$

Where numerator and denominator are independent χ^2.

T

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

Where numerator and denominator are independent.

Transformations If g 1:1 with continuous derivatives and nonzero Jacobian, and $Y = g(X)$, then the density

$$f_Y(y) = f_X(g^{-1}(y))|J_{g^{-1}}(y)|$$

For affine transformation $Y = AX + c$ then

$$f_Y(y) = f_X(A^{-1}(y-c))|\det A|^{-1}$$

Moment generating functions determine distribution

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

X_i independently distributed $\iff$

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

Characteristic function

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

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

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

TODO - add measure theory

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

$$\mathrm{E}\, g(X) \geq g(\mathrm{E}\, X)$$

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

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

$T(X)$ 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

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

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

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

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

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

Also holds for more general convex loss function L:

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

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

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

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

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 $\mathrm{E}\, h(T(x))$ 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.

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

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

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. $\mathrm{rank}(X) = p$ 3. model is identifiable. More generally another equivalent statement is $\mathrm{Var}(T(X)) = \ddot{A}(\eta)$ is positive definite.

Decision Theory

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

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

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

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

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

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

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 $r_{\pi_k}(\delta_{\pi_k}) \to c$.

Asymptotics

Almost Sure convergence

$$X_n \xrightarrow{a.s.} X \quad \text{means} \quad P(X_n \to X) = 1$$

Theorem: $\iff P(\sup_{m \geq n} |X_m - X| > \epsilon) \to 0 \quad \forall \epsilon > 0.$

Convergence in Probability

$$X_n \xrightarrow{p} X$$

means that $P(|X_n - X| > \epsilon) \to 0$ for all $\epsilon > 0$.

Generalized Chebychev Inequality Let X be a r.v. and g be a nonnegative function increasing on the range of X. Then

$$P(X \geq a) \leq \frac{\mathrm{E}\, g(X)}{g(a)}$$

Borel Cantelli Lemma

Hoeffding Inequality Let $X_1, \ldots, X_n$ be independent (not necessarily iid) with $a_i \leq X_i \leq b_i$ and $\mathrm{E}\, X_i = 0$. Then

$$P\left(\sum_{i=1}^{n} X_i \geq \eta\right) \leq \exp\left\{-\frac{2\eta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\}$$

Multivariate Normal

log likelihood for k vector $x \sim N(\mu, \Sigma)$

$$l_x = -\frac{k}{2}\log 2\pi - \frac{1}{2}\{\log\det\Sigma + (x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

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

$$\mathrm{E}(g(X)(X-\mu)) = \sigma^2\,\mathrm{E}(g'(X))$$

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:

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

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:

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

General Techniques

Singular Value Decompostion (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.

$$P = P^T \qquad P^2 = P$$

Covariance of linear transformations

$$Cov(Ay, Bx) = ACov(y, x)B^T$$

Invert 2×2 matrix: $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$

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

Block matrix: $\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} =$

$$\begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

Sum identities:

$$\sum_{k=0}^{\infty} p^k = \frac{p}{1-p} \qquad \sum_{k=0}^{\infty} kp^k = \frac{p}{(1-p)^2} \qquad |p| < 1$$

Integration by parts:

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

Matrix / Vector differentiation

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

$\frac{\partial}{\partial \theta_i} \log(|A|) = tr(A^{-1} \frac{\partial A}{\partial \theta_i})$