1. Estimation & Asymptotics

Properties of Estimators

- **MLE ($\hat{\theta}_{MLE}$):** Argmax of $L(\theta)$. Invariant: $\widehat{g(\theta)} = g(\hat{\theta})$.
- **Consistency:** $\hat{\theta}_n \xrightarrow{P} \theta$. (WLLN implies consistency).
- **Bias:** $B(\hat{\theta}) = E[\hat{\theta}] - \theta$.
- **MSE:** $E[(\hat{\theta}-\theta)^2] = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2$.
- **Fisher Info:** $I_n(\theta) = nI_1(\theta) = -E[\frac{\partial^2}{\partial\theta^2}\ln f(X|\theta)]$.
- **CRLB:** $Var(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$. (If eff, Var = CRLB).

Asymptotic Distributions (Large n)

- **CLT:** $\sqrt{n}(\bar{X}-\mu) \xrightarrow{d} N(0,\sigma^2)$.
- **Delta Method:** If $\sqrt{n}(Y_n-\theta) \xrightarrow{d} N(0,\sigma^2)$, then $\sqrt{n}(g(Y_n)-g(\theta)) \xrightarrow{d} N(0,[g'(\theta)]^2\sigma^2)$.
- **MLE Asymptotics:** $\hat{\theta}_{MLE} \approx N(\theta, \frac{1}{I_n(\theta)})$.
- **Wald Statistic:** $\frac{\hat{\theta}-\theta_0}{\widehat{SE}} \sim N(0,1)$.
- **Score Test:** Using $\frac{\partial}{\partial\theta}\ln L(\theta)\big|_{\theta_0}$.

Inequalities

- **Markov:** $P(|X| \geq a) \leq \frac{E|X|}{a}$ (for $a>0$).
- **Chebyshev:** $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$.

2. Decision Theory

Risk Function: $R(\theta,\delta) = E_\theta[L(\theta,\delta(X))]$. **Bayes Risk**: $r(\pi) = \int R(\theta,\delta)\pi(\theta)d\theta = E[L(\theta,a)|X]$.

Loss Functions Bayes Estimators Minimize posterior expected loss: $\min_a \int L(\theta,a)\pi(\theta|x)d\theta$.

1. **Squared Error Loss** $L(\theta,a) = (\theta-a)^2$
   - Minimize $E[(\theta-a)^2|X]$.
   - $\frac{d}{da}E[\theta^2 - 2a\theta + a^2|X] = -2E[\theta|X] + 2a = 0$.
   - $\hat{\theta}_{Bayes} = E[\theta|X]$ (Posterior Mean).
2. **Absolute Error Loss** $L(\theta,a) = |\theta-a|$
   - Minimize $\int|\theta-a|\pi(\theta|x)d\theta$.
   - $\hat{\theta}_{Bayes} =$ Median of $\pi(\theta|X)$.
   - Satisfies $\int_{-\infty}^{a}\pi(\theta|x)d\theta = 0.5$.
3. **0-1 Loss** (Hit or Miss)
   - $L=0$ if $|\theta-a|<\epsilon$, 1 otherwise. Limit $\epsilon \to 0$.
   - Maximize posterior probability.
   - $\hat{\theta}_{Bayes} = \arg\max_\theta \pi(\theta|X)$ (Posterior Mode / MAP).

3. Hypothesis Testing

Setup: $H_0: \theta \in \Theta_0$ vs $H_1: \theta \notin \Theta_0$. **Type I Error** ($\alpha$): P(Reject $H_0|H_0$ True). **Type II Error** (β): P(Accept $H_0|H_1$ True). **Power**: $1-\beta = P$(Reject $H_0|H_1$ True).

Likelihood Ratio Test (LRT)

$$\Lambda(x) = \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta}_{MLE})}$$

Rejection Region: $\{x : \Lambda(x) \leq c\}$. **Wilks' Thm:** $-2\ln\Lambda \xrightarrow{d} \chi^2_v$, $v = \dim(\Theta) - \dim(\Theta_0)$.

Tests for Mean (Normal)

- **Z-Test (σ known):** $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$. RR: $|Z| > z_{\alpha/2}$.
- **T-Test (σ unknown):** $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$.

Tests for Variance (Normal)

- **Statistic:** $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$.
- **RR (2-sided):** $\chi^2 < \chi^2_{1-\alpha/2}$ or $\chi^2 > \chi^2_{\alpha/2}$.

Categorical Data (Chi-Square)

- **Goodness of Fit:** $Q = \sum\frac{(O_i-E_i)^2}{E_i} \sim \chi^2_{k-1-p}$.
- p: number of estimated parameters.
- O_i: Observed count, E_i: Expected under H_0.

4. Parametric Families

Bernoulli / Binomial $X \sim Bin(n,p)$, $f(x) = \binom{n}{x}p^x(1-p)^{n-x}$.

- **MLE:** $\hat{p} = X/n$ (for Bin), $\bar{X}$ (for Bern).
- **Info:** $I(p) = \frac{n}{p(1-p)}$.
- **Conjugate Prior:** $p \sim Beta(\alpha,\beta) \propto p^{\alpha-1}(1-p)^{\beta-1}$.
- **Posterior:** $p|x \sim Beta(\alpha+x, \beta+n-x)$.
- **Bayes Mean (SE):** $\hat{p}_B = \frac{\alpha+x}{\alpha+\beta+n} = w\frac{\alpha}{\alpha+\beta} + (1-w)\frac{x}{n}$.
- **MAP (Mode):** $\frac{\alpha+x-1}{\alpha+\beta+n-2}$.
- **Jeffreys Prior:** $Beta(1/2,1/2) \propto p^{-1/2}(1-p)^{-1/2}$ (Arcsine).
- **Wald CI:** $\hat{p} \pm z\sqrt{\hat{p}(1-\hat{p})/n}$.
- **Predictive (m future trials):** Beta-Binomial.

$$P(Y=y|x) = \binom{m}{y}\frac{B(\alpha+x+y, \beta+n-x+m-y)}{B(\alpha+x,\beta+n-x)}$$

For 1 trial: $P(X_{new}=1|Data) = E[p|Data] = \hat{p}_B$.

Poisson $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$, $\lambda > 0$.

- **MLE:** $\hat{\lambda} = \bar{X}$.
- **Info:** $I(\lambda) = \frac{n}{\lambda}$.
- **Asymptotics:** $\sqrt{n}(\hat{\lambda}-\lambda) \xrightarrow{d} N(0,\lambda)$.
- **VS Transform:** $\sqrt{\hat{\lambda}}$ stabilizes variance.
- **Conjugate Prior:** $\lambda \sim Gamma(\alpha,\beta) \propto \lambda^{\alpha-1}e^{-\beta\lambda}$.
- **Posterior:** $\lambda|x \sim Gamma(\alpha+\sum x_i, \beta+n)$.
- **Estimators:** Mean $\frac{\alpha+\sum x_i}{\beta+n}$, Mode $\frac{\alpha+\sum x_i - 1}{\beta+n}$.
- **Jeffreys Prior:** $\pi(\lambda) \propto \sqrt{1/\lambda} = \lambda^{-1/2}$ (Improper).
- **Predictive:** Negative Binomial.

$$P(y|x) = \binom{\alpha'+y-1}{y}\left(\frac{\beta'}{\beta'+1}\right)^{\alpha'}\left(\frac{1}{\beta'+1}\right)^{y}$$

where $\alpha' = \alpha + \sum x$, $\beta' = \beta + n$.

Exponential $f(x) = \lambda e^{-\lambda x}$. Mean $1/\lambda$, Var $1/\lambda^2$.

- **MLE:** $\hat{\lambda} = 1/\bar{X}$. (Biased).
- **Info:** $I(\lambda) = n/\lambda^2$.
- **Conjugate Prior:** $\lambda \sim Gamma(\alpha,\beta)$.
- **Posterior:** $\lambda|x \sim Gamma(\alpha+n, \beta+\sum x_i)$.
- **Bayes Estimators:** Mean $\frac{\alpha+n}{\beta+\sum x_i}$.
- **Jeffreys Prior:** $\pi(\lambda) \propto 1/\lambda$ (Improper).
- **Predictive:** Lomax (Pareto Type II).

$$p(x_{new}|x) = \frac{\alpha'(\beta')^{\alpha'}}{(\beta'+x_{new})^{\alpha'+1}}$$

where $\alpha' = \alpha+n$, $\beta' = \beta + \sum x_i$.

- **Relationship:** $2\lambda\sum X_i \sim \chi^2_{2n}$.

Geometric $f(x|p) = (1-p)^{x-1}p$. Mean $1/p$.

- **MLE:** $\hat{p} = 1/\bar{X}$.
- **Info:** $I(p) = \frac{n}{p^2(1-p)}$.
- **Conjugate Prior:** $p \sim Beta(\alpha,\beta)$.
- **Posterior:** $p|x \sim Beta(\alpha+n, \beta+\sum x_i - n)$.
- **Estimators:** Mean $\frac{\alpha+n}{\alpha+\beta+\sum x_i}$.
- **Jeffreys:** $\pi(p) \propto p^{-1}(1-p)^{-1/2}$.

Normal (σ^2 known) $X \sim N(\mu,\sigma^2)$.

- **MLE:** $\hat{\mu} = \bar{X}$.
- **Conjugate Prior:** $\mu \sim N(\mu_0,\tau_0^2)$.
- **Posterior:** $\mu|x \sim N(\mu_n,\tau_n^2)$.

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \quad \text{(Precisions add)}$$

$$\mu_n = \tau_n^2\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{x}}{\sigma^2}\right)$$

- **Predictive:** $X_{new} \sim N(\mu_n, \sigma^2+\tau_n^2)$.
- **Jeffreys:** $\pi(\mu) \propto 1$.

Normal (μ known, σ^2 unknown)

- **Prior:** $\sigma^2 \sim IG(\alpha,\beta)$ or Precision $\tau \sim Gamma(\alpha,\beta)$.
- **Jeffreys:** $\pi(\sigma^2) \propto 1/\sigma^2$.

Uniform $U(0,\theta)$ $f(x) = 1/\theta, 0<x<\theta$.

- **MLE:** $\hat{\theta} = X_{(n)} = \max(X_i)$.
- **Conjugate Prior:** Pareto $P(\theta) \propto \theta^{-(\alpha+1)}I(\theta > x_m)$.
- **Posterior:** $P(\theta|x) \propto \theta^{-(\alpha+n+1)}I(\theta > \max(x_m, X_{(n)}))$.
- This is $Pareto(\alpha+n, \max(x_m, X_{(n)}))$.
- **Bayes Estimator (MSE):** $\frac{K}{K-1}m'$ where $K = \alpha+n, m' = \max$.

$$\hat{\theta}_{Bayes} = \frac{\alpha+n}{\alpha+n-1}\max(x_m, X_{(n)})$$

Multinomial $P(X) \propto \prod\theta_k^{x_k}$.

- **Conjugate Prior:** Dirichlet$(\alpha_1,...,\alpha_k)$.

$$\pi(\theta) \propto \prod\theta_k^{\alpha_k-1}$$

- **Posterior:** Dirichlet$(\alpha_1+x_1,...,\alpha_k+x_k)$.
- **Posterior Mean:** $E[\theta_k|X] = \frac{\alpha_k+x_k}{\sum(\alpha_j+x_j)}$.
- **Marginal Posterior:** $\theta_k|X \sim Beta(\alpha_k+x_k, \alpha_{rest}+x_{rest})$.
- **Jeffreys Prior:** Dirichlet$(1/2,...,1/2)$.

5. Formulas & Proofs

Gamma Func: $\Gamma(n) = (n-1)!$, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.
Beta Func: $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.
Posterior Predictive Integral:

$$p(x_{new}|x) = \int p(x_{new}|\theta)\pi(\theta|x)d\theta$$

Beta-Binomial Mean: $n\frac{\alpha}{\alpha+\beta}$.
Pareto Mean: $\frac{\alpha x_m}{\alpha-1}$ (requires $\alpha>1$).
Mean of Gamma(α,β): α/β.
Var of Gamma: α/β^2.