- Preliminary

↳ Cumulative distribution function (CDF)

$F_X : \mathbb{R} \mapsto [0,1]$, $F_X(x) = P(X \le x)$

↳ $\lim_{x \to \infty} F_X(x) = 1$; $\lim_{x \to -\infty} F_X(x) = 0$

↳ Non decreasing ; $P(a < X \le b) = F_X(b) - F_X(a)$

↳ $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

↳ Probability density function (PDF)

↳ $P(a \le X \le b) = \int_a^b f_X(t)\,dt$

↳ $f_X(x) = F_X'(x)$, non negative

↳ Expected value (E)

↳ $E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$, mean ⇔ g = 1

↳ Linear : $E(aX + bY) = aE(X) + bE(Y)$

↳ Jensen's inequality for convex functions

$\varphi(E(X)) \le E(\varphi(X))$

↳ Variance (σ^2)

↳ $Var(X) = E((X - E(X))^2) = E(X^2) - E^2(X)$

↳ $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

↳ $Cov(X,Y) = E(XY) - E(X)E(Y)$

↳ Cauchy-Schwartz : $Var(X)Var(Y) \ge Cov(X,Y)$

↳ X,Y independentes ⇒ $Cov(X,Y) = 0$

↳ Distributions :

X	PDF	E(X)	Var(X)
Bernoulli(p)	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial(n,p)	$\binom{n}{k} p^k (1-p)^{n-k}$, $k \in \mathbb{N}$	np	$np(1-p)$
Geometric(p) attempts } fail	$p(1-p)^{k-1} \} k$, $k \in \mathbb{N}^+ \} \mathbb{N}$	$\frac{1}{p} \} \frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	$\frac{e^{-\lambda}\lambda^k}{k!}$, $k \in \mathbb{N}$, $\lambda \in \mathbb{R}^+$	λ	λ
Uniform(a,b)	$\frac{1}{b-a}$, $x \in [a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal(μ,σ²)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2
Exponential(λ)	$\lambda e^{-\lambda x}$, $x \in \mathbb{R}_0^+$, $\lambda \in \mathbb{R}^+$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Beta(a,b)	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$, $x \in [0,1]$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
Gamma(a,λ)	$\frac{1}{\Gamma(a)} \lambda^a x^{a-1} e^{-\lambda x}$, $x \in \mathbb{R}_0^+, \lambda \in \mathbb{R}^+$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$

$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\,dt$, $z > 0$ $\begin{cases} \Gamma(1) = 1 \\ \Gamma(z+1) = z\Gamma(z) \end{cases}$

Introduction

· Parametric model : set $P = \{P_\theta, \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^K$, $K > 1$

↳ Identifiable : $\theta \mapsto P_\theta$ is a bijection

↳ Dominated : measure ν st P_θ has a density p_θ with respect to ν, $\forall \theta \in \Theta$

↳ n iid observations $x = (x_1, \dots, x_n)$, $p_\theta(x) = p_\theta(x_1) \cdots p_\theta(x_n)$

· Decision function : δ mapping from observations X to actions A

· Loss function : mapping from $\Theta \times A$ to $\mathbb{R}_+$ st $L(\theta, a)$ is the cost of selecting a for θ

↳ Estimation : $(\theta - a)^2$; CR : $\mathbb{1}\{\theta \ne a\}$

· Risk : expectation of L over all observations

$R(\theta, \delta) = E_\theta(L(\theta, \delta(X)))$

Estimation

· Maximum likelihood (MLE)

↳ $\hat{\theta}(x) = \arg\max_{\theta \in \Theta} p_\theta(x)$

↳ For n iid observations :

$\hat{\theta}(x) = \arg\max_{\theta \in \Theta} \log p_\theta(x) = \arg\max_{\theta \in \Theta} \sum_{i=1}^n \log p_\theta(x_i)$

↳ $l_\theta(x) = \log p_\theta(x)$

↳ Argmax : check 1st derivative = 0

. Method of moments

↳ Strong Law of Large Numbers iid : $E_\theta(X) \approx \frac{1}{n}\sum_{i=1}^n x_i$

↳ $\hat{\theta}(x) = F^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i\right)$, $F : \theta \mapsto E_\theta(X)$

· Bias : $b(\theta, \hat{\theta}) = E_\theta(\hat{\theta}(X)) - \theta$

Quadratic Risk

· Quadratic Risk : $R(\theta, \hat{\theta}) = E((\hat{\theta}(X) - \theta)^2)$

↳ $R(\theta, \hat{\theta}) = b^2(\theta, \hat{\theta}) + Var(\hat{\theta}(X))$

· Fisher information : $I(\theta) = -E\left(\frac{\partial^2 \log p_\theta}{\partial \theta^2}(X)\right)$

↳ Score : $\frac{\partial \log p_\theta(X)}{\partial \theta}$; $E = 0$, $var = I(\theta)$

↳ $I_n(\theta) = n I_1(\theta)$, n iid

· Cramér-Rao Bound ; $R(\theta, \hat{\theta}) \ge I^{-1}(\theta)$

↳ Unbiased and $R(\theta, \hat{\theta}) = I^{-1}(\theta)$ ⇒ efficient

· General case

↳ $b(\theta, \hat{g}) = E(\hat{g}(X)) - g(\theta)$

↳ $R(\theta, \hat{g}) = E((\hat{g}(X) - g(\theta))^2)$

↳ $R(\theta, \hat{g}) = b^2(\theta, \hat{g}) + var(\hat{g}(X))$

↳ $R(\theta, \hat{g}) \ge g'^2(\theta) I^{-1}(\theta)$

· Vectorial case

↳ $I(\theta) = -E\left(\frac{\partial^2 \log p_\theta(X)}{\partial \theta_i \partial \theta_j}\right)_{i,j = 1, \dots, K}$

↳ $I(\theta) = cov(\nabla \log p_\theta(X))$

↳ $R(\theta, \hat{g}) \ge \nabla g(\theta)^T I(\theta) g(\theta)$

Bayesian Statistics

· Prior distribution $\pi(\theta)$ before any obs.

· Posterior distribution $\pi(\theta | x) = \frac{\pi(\theta) p(x|\theta)}{p(x)} \propto \pi(\theta) p(x|\theta)$

· Bayes estimator : $\hat{\theta}(x) = E(\theta | x)$

· Bayes risk : $r(\hat{\theta}) = E(R(\theta, \hat{\theta}))$

↳ $r(\hat{\theta}) = E(E((\hat{\theta}(X) - \theta)^2 | X))$, min for $\hat{\theta}(x) = E(\theta | x)$

· Bayes bias : $b(\theta, \hat{\theta}) = E(\hat{\theta}(X) | \theta) - \theta$

↳ $R(\theta, \hat{\theta}) = b^2(\theta, \hat{\theta}) + Var(\hat{\theta}(X) | \theta)$

· Conjugate prior : prior that belongs to the same family as the posterior

↳ Bernoulli - Beta

↳ Binomial - Beta

↳ Poisson - Gamma

↳ Multinomial - Dirichlet

↳ Gaussian (mean unknown) - Gaussian

↳ Gaussian (variance unknown) - Inverse Gamma

↳ Exponential - Gamma

· Exponential family

↳ $p_\theta(x) = h(x) \exp(\eta(\theta)^T T(x) - A(\theta))$

↳ $\pi(\theta) \propto \exp(\eta(\theta)^T \alpha + \beta A(\theta))$

· Jeffreys prior : $\pi(\theta) \propto \sqrt{I(\theta)}$

↳ Not informative

Hypothesis Testing

· Null hypothesis H_0 considered as true in the absence of any obs. (default) & alternative hypothesis H_1

· Decision : $\delta(x) = \begin{cases} 0 \to \text{accept } H_0 \\ 1 \to \text{reject } H_0 \text{ in favor of } H_1 \end{cases}$

· Errors

↳ Type I (false positive) → $\alpha = P(\delta(X) = 1 | H_0)$

↳ Type II (false negative) → $\beta = P(\delta(X) = 0 | H_1)$

· Neyman-Pearson ; focus on α

↳ Level : $\alpha = \sup_{\theta \in \Theta_0} P_\theta(\delta(X) = 1 | H_0)$

↳ Power : $1 - \beta(\theta)$; $\forall \theta \in \Theta_1$, $\beta(\theta) = P_\theta(\delta(X) = 0)$

↳ p-value is the probability of sampling a value at least as extreme as the true observation under H_0

· Parametric models

↳ $H_0 \to \Theta_0 \subset \Theta$, $H_1 \to \Theta_1 \subset \Theta$

↳ Uniformly most powerful (UMP) such that $\sup_{\theta \in \Theta_0} P_\theta(\delta'(X) = 1) \le \alpha \Rightarrow \forall \theta \in \Theta_1, P_\theta(\delta'(X) = 0) \ge \beta$

· Simple hypothesis

$\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$

↳ UMP : $\delta(x) = \mathbb{1}\left\{\frac{p_1(x)}{p_0(x)} > c\right\}$, $\forall c > 0$

$\alpha = P_0(\delta(X) = 1) = P_0\left(\frac{p_1(X)}{p_0(X)} > c\right)$

· One tailed hypothesis

$\Theta_0 = \{\theta \le \theta_0\}$, $\Theta_1 = \{\theta > \theta_0\}$

UMP $\forall c > 0$

$\frac{p_{\theta'}(x)}{p_\theta(x)} = F(T(x))\ \forall \theta' > \theta$; $\delta(x) = \begin{cases} \mathbb{1}(T(x) > c), & F\uparrow \\ \mathbb{1}(T(x) < c), & F\downarrow \end{cases}$

$\alpha = P_0(\delta(x) = 1) = \begin{cases} P_0(T(X) > c), & F\uparrow \\ P_0(T(X) < c), & F\downarrow \end{cases}$

- Tests tailed

$\Theta_0 = [\theta_1, \theta_2]$, $\theta_1 < \theta_2$, $\Theta_1 = \mathbb{R} \setminus \Theta_0$

$\alpha = P_0(T(x) < c_1) + P_0(T(x) > c_2)$

- Unbiased test: $P_\theta(\delta(x) = 1) \geq \alpha$, $\forall \theta \in \Theta$

Bayesian tests

- Bayes risk: $r(\delta) = E(R(\theta, \delta)) =$
$= \int_{\Theta_0} P_\theta(\delta(x) = 1)\pi(\theta)\, d\mu(\theta) +$
$\int_{\Theta_1} P_\theta(\delta(x) = 0)\pi(\theta)\, d\mu(\theta)$

- Bayesian test: $\delta(x) = \mathbb{1}\left\{\frac{\pi(\Theta_1|x)}{\pi(\Theta_0|x)} > 1\right\}$

↳ $r(\delta) = \pi(\Theta_0)\alpha + \pi(\Theta_1)\beta$

χ^2 tests

- Test for fit

$H_0 = \{X \sim p\}$, $H_1 = \{X \nsim p\}$

$p = \{p_1, \dots, p_K\}$, n iid $X_1, \dots, X_n$

↳ Counts for each category $N_i = \sum_{t=1}^n \mathbb{1}\{X_t = i\}$

↳ $E = np_i$, $var = np_i(1-p_i)$

↳ Statistic: $T(x) = \sum_{i=1}^K \frac{(N_i - np_i)^2}{np_i} \xrightarrow[n \to \infty]{d} \chi^2(K-1)$

↳ $\delta(x) = \mathbb{1}\{T(x) > c\}$, $\alpha = P_0(T(X) > c)$

↳ $np_K \geq 5$

- Test for independence

$H_0 = \{X \perp Y\}$, $H_1 = \{X \not\perp Y\}$

$A_1, \dots, A_K$; $B_1, \dots, B_L$; n iid $(X_1, Y_1), \dots, (X_n, Y_n)$

↳ $N_{ij} = \sum_{t=1}^n \mathbb{1}\{X_t \in A_i, Y_t \in B_j\}$

↳ $N_i = \sum_{t=1}^n \mathbb{1}\{X_t \in A_i\}$, $N_j = \sum_{t=1}^n \mathbb{1}\{Y_t \in B_j\}$

↳ Statistic: $T(X,Y) = \sum_{ij} \frac{\left(N_{ij} - \frac{N_i N_j}{n}\right)^2}{\frac{N_i N_j}{n}}$

$\xrightarrow{d} \chi^2((K-1)(L-1))$, $n \to \infty$

↳ $\delta(x,y) = \mathbb{1}\{T(x,y) > c\}$, $\alpha = P_0(T(X,Y) > c)$

Confidence region

- Decision function: $\forall x$, $\delta(x) \subset \Theta$
- δ provides a confidence region at level $1-\alpha$ if $\forall \theta \in \Theta$, $P_\theta(\theta \in \delta(x)) \geq 1-\alpha$

↳ $\theta \in \mathbb{R}$

↳ Interval: $\delta(x) = [m(x), M(x)]$

↳ Lower bound: $\delta(x) = [m(x), +\infty)$

↳ Upper bound: $\delta(x) = (-\infty, M(x)]$

- Pivot function: φ_θ is a pivot function if the distribution of the random variable $\varphi_\theta(X)$ is independent of θ
- Finding a confidence region at level $1-\alpha$ is equivalent to test the null hypothesis $H_0 = \{\theta = \theta_0\}$ against the alternative hypothesis $H_1 = \{\theta \neq \theta_0\}$, at level α for each $\theta_0 \in \Theta$

- Gaussian model with unknown mean and variance

↳ Pivotal function: $Z = \frac{\bar{X} - \mu}{\sqrt{V_0/n}}$, $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$,

$V_0 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$

↳ Z has Student's distribution with $n-1$ df

↳ $St(n) \sim T \sim \frac{X}{\sqrt{\frac{Y}{n}}}$, $X \sim N(0,1)$, $Y \sim \chi^2(n)$

Extra

- Integration by parts:

$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx$

- $e^{\ln x} = x$
- Quantile function: $Q(p) = F_X^{-1}(p)$
- Distributions

X	MLE	b	MME	b
Bernoulli(p)	$\hat{p} = \bar{x}$	0	$\hat{p} = \bar{x}$	0
Binomial(m,p) (unknown p)	$\hat{p} = \frac{\bar{k}}{m}$	0	$\hat{p} = \frac{\bar{k}}{m}$	0
geometric(p) $k \geq 0$	$\hat{p} = \frac{1}{1+\bar{k}}$	$\neq 0$	$\hat{p} = \frac{1}{1+\bar{k}}$	$\neq 0$
geometric(p) $k \geq 1$	$\hat{p} = \frac{1}{\bar{k}}$	$\neq 0$	$\hat{p} = \frac{1}{\bar{k}}$	$\neq 0$
Poisson(λ)	$\hat{\lambda} = \bar{k}$	0		
Uniform(0,b)	$\hat{b} = \max x$		$\hat{b} = 2\bar{x}$	
normal(μ,σ^2) (unknown μ)	$\hat{\mu} = \bar{x}$	0	$\hat{\mu} = \bar{x}$	0
normal(μ,σ^2) (unknown σ^2)	$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$	$\neq 0$	$\sigma^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \mu^2$	
normal(μ,σ^2) (unknown μ,σ^2)	$\hat{\mu} = \bar{x}$, $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \hat{\mu})^2$		" "	
exponential(λ)	$\hat{\lambda} = \frac{1}{\bar{x}}$	$\neq 0$		
Beta(a,1)	$\hat{a} = \frac{-n}{\sum_{i=1}^n \log x_i}$	$\neq 0$	$\hat{a} = \frac{\bar{x}}{1-\bar{x}}$	$\neq 0$
gamma(a,λ) (unknown $\hat{a},\hat{\lambda}$)			$\hat{a} = \frac{\bar{x}^2}{\bar{x}^{(2)} - \bar{x}^2}$, $\hat{\lambda} = \frac{\bar{x}}{\bar{x}^{(2)} - \bar{x}^2}$	

- $\sum_{i=1}^K Z_i^2 \sim \chi^2(K)$, $Z \sim N(0,1)$

↳ $E = K$, $Var = 2K$

- $T \sim St(\nu)$

↳ $E = 0$, $\nu > 1$

↳ $Var = \frac{\nu}{\nu - 2}$, $\nu > 2$, ∞ $\nu \in]1, 2]$