Parametric Families of distributions

↳ class of pmf or pdfs that are indexed by a parameter or set of parameters

Ex $X \to \{0,1\}$

$p \in (0,1)$

Bernoulli (p)

$p(0) = 1-p$

$p(1) = p$

Ex $\lambda > 0$ $\exp(\lambda)$

$$f(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x > 0 \end{cases}$$

μ, σ^2 $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Normal distributions

↳ Central Limit Theorem

$X_1, \ldots, X_n \overset{iid}{\sim}$ $E[X_i] = \mu$

$Var(X_i) = \sigma^2$

$$\bar{X} = \frac{1}{n} \sum X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

for n large

<u>Note</u> $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

for any n, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

<u>Estimator</u> $X_1, \ldots, X_n \sim f(x \mid \theta_1, \ldots, \theta_r)$

estimate some function

$g(\theta_1, \ldots, \theta_r)$

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

$g(\mu, \sigma^2) = \frac{\mu}{\sigma}$ μ , σ^2

biased / unbiased estimators.

$\phi_n(X_1, \dots, X_n)$ estimator $g(\theta_1, \dots, \theta_r)$

$E[\phi_n(X_1, \dots, X_n)] = g(\theta_1, \dots, \theta_r)$

unbiased

if not, it is considered biased

$X_1, \dots X_n \sim \exp(\lambda)$

$\phi_n = \frac{n}{\sum X_i} = \frac{1}{\bar{X}}$

$E[\phi_n] = \frac{n}{n-1} \lambda$ $\rightarrow$ biased

$\phi_n^* = \frac{n-1}{n} \phi_n$ $E[\phi_n^*] = \frac{n-1}{n} \cdot E[\phi_n]$

$= \lambda$

$\rightarrow$ unbiased

bias of estimator $E[\phi_n] - g(\theta_1, \dots, \theta_r)$

Maximum Likelihood Estimation

$X_1, \ldots, X_n \sim f(x \mid \theta_1, \ldots, \theta_r)$

$L(\theta_1, \ldots, \theta_r \mid x_1, \ldots, x_n)$

$$= \prod_{i=1}^{n} f(x_i \mid \theta_1, \ldots, \theta_r)$$

<u>Ex</u> $X_1, \ldots, X_3 \sim \exp(\lambda)$

$x_1 = 1$

$x_2 = 2$

$x_3 = 3$

$$L(\lambda \mid 1, 2, 3) = \lambda^3 \exp(-\lambda(1+2+3))$$

↳ the idea behind maximum likelihood is that we choose the parameter(s) that maximizes the likelihood

$$\log(\underbrace{L}(\overbrace{\theta_1, \ldots, \theta_r}^{fx} \mid x_1, \ldots, x_n))$$

$\ell(\theta_1, \dots, \theta_r \mid X_1, \dots, X_n)$

Ex $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

$$\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$E\left[\hat{\sigma}^2_{MLE}\right] = \frac{n}{n-1} \sigma^2$$

sometimes MLEs have bias

↳ bias goes to zero as $n \to \infty$

MLEs are "consistent", so that

$$P(|\hat{\theta}_{MLE} - \theta| > \varepsilon) \to 0$$

as $n \to \infty$

Method of moments

$X_1, \dots, X_n \sim f(x \mid \theta_1, \dots, \theta_r)$

kth moment of X

$$E[X^k]$$

$\Rightarrow$ find a formula for

$$E[X^k] = g_k(\theta_1, \dots, \theta_r)$$

for $k = 1, \dots, r$

solve for $\hat{\theta}_{1,mm}, \dots, \hat{\theta}_{r,mm}$

$$g_k(\theta_1, \dots, \theta_r) = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

Ex $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

$$E[X_i] = \mu$$

$$E[X_i^2] = \sigma^2 + \mu^2$$

solve $\hat{\mu} = \frac{1}{n} \sum X_i$

$$\sigma^2 + \mu^2 = \frac{1}{n} \sum X_i^2$$

Ex Gamma distribution with parameters

$\underline{k}, \theta$

Gamma function: $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} \, dy$

$n = 1, 2, \dots, \quad \Gamma(n) = (n-1)!$

$X_1, \dots, X_k \sim \exp(1/\theta)$

$X = \sum_{i=1}^{k} X_i \sim \text{Gamma}(k, \theta)$

$$f(x \mid k, \theta) = \begin{cases} \dfrac{x^{k-1}}{\Gamma(k)\,\theta^k} \cdot e^{-x/\theta}, & x > 0 \\ 0, & \text{o/w} \end{cases}$$

$k > 0, \quad \theta > 0$

$$L(k, \theta \mid x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{x_i^{k-1}}{\Gamma(k)\,\theta^k} e^{-x/\theta}$$

$\ell(k, \theta \mid x_1, \dots, x_n)$.

$$= -n \log(\Gamma(k)) - nk \log(\theta) + (k-1) \sum_{i=1}^{n} \log(x_i) - \frac{1}{\theta} \sum_{i=1}^{n} x_i$$

$$\frac{\partial \ell}{\partial \theta} = -\frac{nk}{\theta} + \frac{1}{\theta^2} \sum x_i$$

$$\hat{\theta}_{MLE} = \frac{1}{nk} \sum_{i=1}^{n} x_i$$

calculate an estimator for k

$\Longrightarrow$ numerical routine that calculates $\hat{k}_{MLE}$ and $\hat{\theta}_{MLE}$

Given data $X_1, \ldots, X_n$ from $f(x \mid \theta)$

a $(1-\alpha)\cdot 100\%$ confidence interval is a statistic that takes the form (L, U)

s/t $P(\theta \in (L, U)) = 1-\alpha$

$S_1 = \{X_{11}, \ldots, X_{1n}\} \rightarrow (L_1, U_1)$

$S_2 = \{X_{21}, \ldots, X_{2n}\} \rightarrow (L_2, U_2)$

$\vdots$

$S_I = \{X_{I1}, \ldots, X_{In}\} \rightarrow (L_I, U_I)$

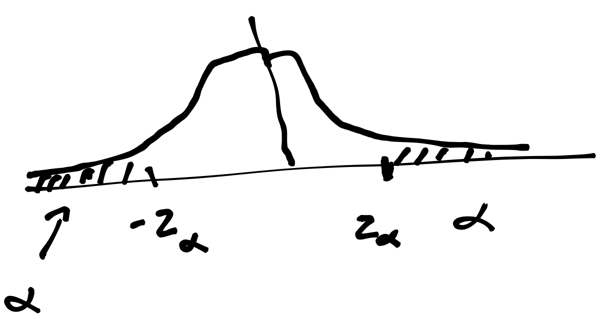
I large

95% of the samples (L_i, U_i) contains θ

$S_1, \dots, S_I$

$Z \sim N(0,1)$

z_α is the value s/t $P(Z \geq z_\alpha) = \alpha$



scipy.stats.norm.ppf$(1-\alpha)$
$= z_\alpha$

$P(|Z| \geq z_{\alpha/2}) = \alpha$