(2) Let $X_1, \ldots X_n$ be i.i.d. $U(0,1)$

Suppose, we are interested in asymptotic distⁿ ($n \to \infty$) of

$$G_n = \left(\prod_{i=1}^{n} X_i\right)^{1/n}$$

Note that $H_n = -\log G_n = \frac{1}{n}\sum_{i=1}^{n}(-\log X_i)$

$$G_n = e^{-H_n}$$

Realize that $X_i \sim U(0,1)$

$$Y_i = -\log X_i \sim \exp(1) \;;\; EY_i = 1,\; VY_i = 1$$

By CLT $\sqrt{n}\left(\bar{Y}_n - 1\right) \xrightarrow{L} N(0,1)$

i.e. $\sqrt{n}\left(H_n - 1\right) \xrightarrow{L} N(0,1)$

Take $g(x) = e^{-x}$ $g'(x) = -e^{-x}$ $g'(1) = -e^{-1} \neq 0$

Δ-rule $\Rightarrow \sqrt{n}\left(e^{-H_n} - e^{-1}\right) \xrightarrow{L} N(0, e^{-2})$

i.e. $\sqrt{n}\left(G_n - e^{-1}\right) \xrightarrow{L} N(0, e^{-2})$

i.e. $\sqrt{n}\left(\left(\prod X_i\right)^{1/n} - e^{-1}\right) \xrightarrow{L} N(0, e^{-2})$

(3) $X_1, \ldots X_n$ i.i.d χ^2_m

$$Y_n = \sum_{i=1}^{n} X_i \sim \chi^2_{nm}$$

Suppose we want to find approximate value of

$P(a < Y_n < b)$ for large n $(a < b)$

By CLT $\dfrac{Y_n - nm}{\sqrt{2nm}} \xrightarrow{L} N(0,1)$

$$P(a < Y_n < b) = P\left(\frac{a - nm}{\sqrt{2nm}} < \frac{Y_n - nm}{\sqrt{2nm}} < \frac{b - nm}{\sqrt{2nm}}\right)$$

$$\underset{\text{for large } n}{\approx} \Phi\left(\frac{b - nm}{\sqrt{2nm}}\right) - \Phi\left(\frac{a - nm}{\sqrt{2nm}}\right)$$

(4) $X_1, \ldots X_n$ i.i.d $P(\lambda)$

$$S_n = \sum_1^n X_i \sim P(n\lambda) \qquad E S_n = n\lambda \quad V S_n = n\lambda$$

By CLT $\quad \dfrac{S_n - n\lambda}{\sqrt{n\lambda}} \xrightarrow{\mathcal{L}} N(0,1)$

Take $n = 64, \ \lambda = 0.125 \ ; \ n\lambda = 8$

$P(S_n = 10) = 0.099 \rightarrow$ from Poisson p.m.f.

Approximate value of the above can be obtained thro CLT

$P(S_n = 10) = P(9.5 < S_n < 10.5) \leftarrow$ continuity correction

$$\underset{\text{for large } n}{\approx} \Phi\left(\frac{10.5 - \overset{n\lambda}{8}}{\sqrt{8}}\right) - \Phi\left(\frac{9.5 - 8}{\sqrt{8}}\right) = .108$$

for $n = 96 \ ; \ \lambda = 0.125, \ n\lambda = 12$

$P(S_n = 10) = .105$ (exact)

$P(S_n = 10) \approx .101$ (thro normal approximation)

Higher the value of n, closer will be the approximation.

Statistical Inference

Let X be a random variable describing a characteristics

We assume an "appropriate" prob model for X, in the sense that we say X has a prob distⁿ with p.d.f. or p.m.f $f_\theta(x)$; where θ is an unknown parameter (or a parameter vector) characterizing the distⁿ of the r.v. X

e.g. $X \sim N(\mu, \sigma^2)$ $\quad \underset{\sim}{\theta} = (\mu, \sigma)'$; $\mu \in \mathbb{R}, \sigma > 0$

$X \sim B(n, \theta)$ $\quad 0 < \theta < 1$

$X \sim \exp(\theta)$, $\quad \theta > 0$

$X \sim P(\theta)$, $\quad \theta > 0$

In each of these cases θ is unknown and is assumed to vary in a space called parameter space, Θ, say

e.g. $X \sim N(\mu, \sigma^2)$; $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$

$X \sim \exp(\theta)$; $\Theta = \{\theta : \theta > 0\}$

Family of distⁿs:

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}.$$

P_θ is a distⁿ characterised by θ.

p.m.f. or p.d.f. of P_θ is $f_\theta(x)$ say

Inference problem: to make inference about the unknown parameter(s).

Random Sample from P_θ :

Let $X_1, \ldots, X_n$ be a random sample from P_θ

i.e. $X_1, \ldots, X_n$ are i.i.d. with p.d.f. (or p.m.f) $f_\theta(x)$.

We use the random sample $(X_1, \ldots, X_n)$ for inference about θ.

3 approaches of statistical inference

(i) **Point estimation**

$\delta(X_1, \ldots, X_n)$ - estimator (f^n of r.v.s); a statistic

$\delta(\underset{\sim}{X}) = \hat{\theta}(\underset{\sim}{X})$; e.g. Maximum likelihood estimator (MLE)

~~For~~

Bayes estimator, unbiased estimator

Uniformly minimum variance unbiased estimator (UMVUE),

Least Squares estimator (LSE)

- - . .

For an observed sample

$(x_1, \ldots, x_n) \longrightarrow \delta(\underset{\sim}{x})$; an estimate

(ii) **Interval estimation** : Confidence interval

Construct a random interval

$[\hat{\theta}_L(\underset{\sim}{X}), \hat{\theta}_U(\underset{\sim}{X})] \ni$

$P\left(\theta \in [\hat{\theta}_L(\underset{\sim}{X}), \hat{\theta}_U(\underset{\sim}{X})]\right) \geq 1 - \alpha$; say $\alpha = 0.05$ or 0.01

Estimated interval from observed sample $(x_1, \ldots, x_n)$

$[\hat{\theta}_L(\underset{\sim}{x}), \hat{\theta}_U(\underset{\sim}{x})]$

(iii) Hypothesis testing

Validate / test some prior belief about parameter

$H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_1$

null hypothesis

alternate hypothesis

θ_0 & θ_1 known consts

or $H_0 : \theta = \theta_0$ against $H_A : \theta > \theta_0$ or $H_A : \theta < \theta_0$

or $H_0 : \theta = \theta_0$ against $H_A : \theta \neq \theta_0$.