Note 2

Haojie Liu

2023-10-03

Possion

Continue from last class:

$$\lambda = 2$$

the expected number of calls is 2 per minute, and the # of calls in X minutes with

$$N_x \sim Possion(\lambda x)$$

P(wait at least X minutes for the first call)

$$P(X > x) = \int_x^\infty f_{exp}(t)dt$$

$$P(X > x) \to P(N_x = 0) = \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = e^{-\lambda x}$$

Question:

Compute that an event does occur during x wait of time

$$P(X > x) \to P(N_x = 1) = \frac{(\lambda x)^1 e^{-\lambda x}}{1!}$$

Beta

Bayesian statistics

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$
$$0 \le x \le 1$$

$$\Gamma(r) = (r-1)!$$
$$\Gamma(1) = 1$$
$$\Gamma(\frac{1}{2}) = \pi^{\frac{1}{2}}$$
$$\alpha = \beta = 1$$

$$Beta(1,1) = 1 \sim Uniform(0,1)$$

Law of Large Number

Weak Law of Large Number

$$lim_{n \to x} P(|\bar{x} - \mu| > \epsilon) = 0$$
$$\bar{X}_n = \frac{\Sigma_{i=1}^{n} X_i}{n}$$

$\bar{X}$ converges in probability to μ

Strong Law of Large Number

$$P(lim_{n \to \infty} |\bar{x} - \mu| > \epsilon) = 0$$

$$\hat{\theta} : estimator \to \theta$$
$$\bar{x} : estimator \to \mu$$
$$s^2 : estimator \to \sigma^2$$

$$\hat{\sigma}^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$$

$$bias : [E(\hat{\theta}) - \theta]^2$$

This is the bias-Variance trade off

$$\hat{p} = \frac{x}{n}$$

$$E(\hat{p}) = p$$

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

$$lim_{n \to \infty} Var(\hat{p}) = 0$$

$\hat{p}$ is a consistant estimator of p when $lim_{n \to \infty} MSE(\hat{\theta}) = 0$

$$x_i \sim F_x(x|\theta) \to (x_1, ...x_n) \to \hat{\theta} = S(x_1, x_n) \to \theta$$

$$\hat{F}_x \to F_x(x)$$

Check with the MSE to see if this is a consistent estimator.

$$F_x(x_0) = P(X \leq x_0)$$

EDF:

$$\hat{F}_x(x_0) = \frac{count(x_1 \leq x_0)}{n} = \frac{\Sigma_{i=1}^{n} I(x_i \leq x_0)}{n}$$

Ex: Given 1, 2, 2, 3, 5; $\hat{F}(1) = \frac{1}{5}; \hat{F}(2) = \frac{3}{5}; \hat{F}(3) = \frac{4}{5}; \hat{F}(5) = 1$

0 if $x < 1$ $x < x_{(1)}$

$\frac{1}{5}$ if $1 \leq x < 2$ $x_{(1)} \leq x < x_{(3)}$

$\frac{3}{5}$ if $2 \leq x < 3$ $x_{(3)} \leq x < x_{(4)}$

$\frac{4}{5}$ if $3 \leq x < 5$ $x_{(4)} \leq x < x_{(5)}$

1 if $5 \leq x$ $x_{(5)} \leq x$

or

$0, x < x_{(1)}$

$\frac{i}{n}, x_{(i)} < x < x_{(i+1)}$

$1, x \geq x_{(n)}$

Population CDF

$$F_x(x_0) = \frac{\Sigma_{i=1}^{N} I(x_i \leq x_0)}{N}$$

$$Set : Y_i = I(X_i \leq x_0)$$

$$Y_i \sim Ber(F_x(x_0))$$

$$\hat{F}_n(x_0) = \frac{\Sigma_{i=1}^{n} I(x_i \leq x_0)}{n}$$

$$E(\hat{F}_n(x_0)) = \frac{n}{n} E(I(x_1 \leq x_0)) = F_x(x_0)$$

$$Var(\hat{F}_n(x_0)) = \frac{1}{n^2} \Sigma_{i=1}^{n} Var(I(x_1 \leq x_0)) = \frac{F_x(x_0)(1 - F_x(x_0))}{n} \leq \frac{1}{4n}$$

$$lim_{n \to \infty} MSE[\hat{F}_n(x_0)] \to 0$$

Therefore $\hat{F}_n(x_0)$ is a consistent estimator

$$x_1, ..., x_{50} \sim Uniform(0, 2)$$

$$E_n(\hat{F}(1))?$$

$$Var(\hat{F}(1))?$$

$$n^{1/2}(\hat{F}_n(x_0) - F(x_0)) \sim N(0, F(x_0)[1 - F(x_0)])$$

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

CI estimator $\pm Z_{1-\alpha} SD$