

Lecture Notes 11

October 25, 2015

Scribe: Wei-Chang Lee, Chi-Ning Chou

0.1 Discrete Distributions

Definition 1 (Discrete Uniform Distribution) Suppose X follows discrete uniform distribution then it has density

$$f_X(|N) = \frac{1}{N} \mathbb{1}_{\{1,2,3,\dots,N\}}(x)$$

where N is an integer, with notation $X \sim \text{Discrete Uniform}(N)$.

Property 1 Given N , X follows discrete uniform distribution then,

1. $E[X|N] = \sum_{i=1}^N P(X=i) i = \frac{N+1}{2}$
2. $\text{Var}(X|N) = E[X^2|N] - E[X|N]^2 = \frac{N^2-1}{12}$

Intuition (Usage in statistics)

How can we test the two given data group X, Y follows the same distribution?

$X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F_1(x)$ and $Y_1, Y_2, \dots, Y_m \stackrel{iid}{\sim} F_2(x)$ want to test $H_0 : F_1(x) = F_2(x) \forall x$

Kolmogorov statistics: Using empirical distribution of $F_1(x), F_2(x)$

$$\hat{F}_1(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x)$$

$$\hat{F}_2(x) = \frac{1}{m} \sum_{i=1}^m \mathbb{1}(Y_i \leq x)$$

We have Kolmogorov statistics:

$$\sup_x |\hat{F}_1(x) - \hat{F}_2(x)|$$

which can not be too big if X and Y following same distributions.

Rank statistics (Wilcoxon test): Instead of using true value as the predictor. We use the order i.e. rank of the data in the group. We combine X and Y and sort them to give rank

$$W = \frac{1}{n} \sum_{i=1}^n \text{Rank}(X_i)$$

To prevent the issue that extreme values dominated the statistics. And $X \sim Y$ if W is not too big or too small.

Definition 2 (Bernoulli Distribution) X follows Bernoulli distribution then it has density

$$f_X(x|p) = p^x(1-p)^{1-x} \mathbb{1}_{\{0,1\}}(x)$$

where $0 \leq p \leq 1$ denoting as $X \sim \text{Bernoulli}(p)$.

Property 2 Given p , X follows binomial distribution then,

1. $E[X^m|p] = E[X|p] = p$
2. $\text{Var}(X|p) = p - p^2 = p(1-p)$
3. $F_X(x) = P(X \leq x) = E[I(X \leq x)] = E[N(x)]$

Definition 3 (Binomial Distribution) X_1, X_2, \dots, X_n i.i.d follows Bernoulli(p), let $X = \sum_{i=1}^n X_i$, X follows binomial distribution having density

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \mathbb{1}_{\{0,1,2,3,\dots,n\}}(x)$$

Intuition (Independent)

X_1, X_2, \dots, X_n are independent iff

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n f(x_i|p)$$

The mutually independent property automatically satisfied since we can think of $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$ where $\Omega_i = \{0,1\}$ for the i th Bernoulli trial. And $\bigcup (A_i \in \Omega_1)(\Omega_2 \times \dots \times \Omega_n)$ augmented Ω .

Remark 1 In reality, X_1, X_2, \dots, X_n are not i.i.d.. Since we sometimes sample population with common factors. They may affect each other, within positive or negative relation.

1. $\rho > 0$: over-dispersion binomial distribution.
2. $\rho < 0$: under-dispersion binomial distribution.