Descriptive statistics

Thursday, 3 February 2022 11:45

Density & distribution functions

Cumulative distribution function (C.D.F)

(distribution function) F

$$F(x) = P(X \leq x) = P(\omega \in \Omega : X(\omega) \leq x)$$

when $X = \{x_1, x_2, \dots\}$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X \leq x_i)$$

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) =$$

For a continuous case.

$$F(x) = \int_{-\infty}^{x} f(y)\,dy.$$

If we have several random variables.

Sample space Ω: exoplanets within 50f

random variables could be $\{X_1, X_2, X_3\}$

X_1: exoplanet mass

X_2 = radii

X_3 = surface temperature

X_4 = orbital axis

Vector of random variable: $(X_1, X$

$$F(x_1 \ldots x_k) = P(X_1 \le x_1, \ldots, X_k \le x_k)$$

$$= P(\omega \in \Omega : X_1(\omega) \le x_1 \ldots X_k(\omega)$$

Marginal distribution:

For an individual X_i

$$\boxed{F(x_i) = P(X_i \le x_i)}$$

Expected Means

The Mean of a random variable is defined weighted average where the weight is ob from the associated probabilites

$$E(X) = \sum x_i \, P(X = x_i) =$$

if $Y = h(x)$: real-valued function

$$E(h(x)) = \sum_i h(x_i) P(X = x_i)$$

$E[W] = ?$

$E[W]$ cannot de defined.

$W^+ = \max(0, W)$

$W^- = \max(0, -W)$

$$\sum_i |x_i| P(X = x_i) < \infty$$

finite

$E[W] =$

Which is

Continuous Case:

$$E(X) = \int_{-\infty}^{\infty} y f(y) dy$$ Provided

$$E[h(X)] = \int_{-\infty}^{+\infty} h(y) f(y) dy \quad " \quad \int_{-\infty}^{+\infty} |h$$

Variance:

Quantifies the "Spread" is the S
Centered on the Mean.

$$\sigma^2 = Var(X) = E[(X-\mu)^2] -$$

$$\mu = E[X]$$

If we have $X_1, X_2, \dots X_n$.

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

$$Var\left(\sum X_i\right) = \sum Var(X_i) + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ i \neq j}}^{N} Cov($$

$$Cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

Covariance b/w X & Y.

Measures the relationship b/w th

in two random variables.

If all the X_i variables have the same Va

the situation is called "homoscedastic

the variance of the sample mean $\bar{X} =$

$$Var(\bar{x}) = \frac{1}{n^2} \sum Var(x_i) = \sigma^2/n.$$

Standard deviation $= \sigma = \sqrt{V}$

Standardized form:

$$\boxed{X_{std} = \frac{X - \mu}{\sigma}}$$

- Normalisation. — Zero mean and u
- logarithmic transformation.

Skewness

$$E[(X - E[X])^3] \rightarrow Thi$$

Quantile functions

$F(X) \rightarrow$ estimates the value

Population distribution function

value of x.

Astronomy: What value of x co... to a specified value ...

What fraction of galaxies hav... above L^*?

threshold value.

$$Q(u) = F^{-1}(u) = \inf\{y : F ...$$

where $0<u<1$. inf: infimum ... value of y with the property sp... the brackets.

quantiles such as 5%, 25%, 50... 95%

Quantile-Quantile Plots (Q...

Next lecture:

* Estimators
* Sample Mean, Sampl
* Variance / Mean of th
* Real Data;
* outliers.
* Distribution functions