Set Theory

15 October 2017 23:47

- ω $\in A$
  w belongs to A
- <u>Subset</u> : $A \subseteq B$
  - all elements of A are contained in set B

<u>Empty set</u>: $\emptyset = \{\ \}$

<u>Universal set</u>: Ω
all elements

Cardinality - number of elements in a set
Singleton - cardinality 1

De Morgan's laws

break the bar
change the operator

- $\overline{A \cup B} = \bar{A} \cap \bar{B}$
- $\overline{A \cup B} = \bar{A} \cap \bar{B}$

Disjoint: no shared elements

Partition: collection of disjoint sets form Ω

<u>Distributive Law</u>
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

<u>Union:</u> $A \cup B$
all elements that belong to at least one of the sets

<u>Intersection:</u> $A \cap B$
all elements that belong to all of the sets

<u>Complement:</u> not A
$\bar{A} =$ all elements in Ω not in A

<u>Difference:</u> A/B
elements in A that are also not in B
$A/B = A \cap \bar{B}$

Sample Spaces and Events

28 October 2017 18:07

S - universal set
Event - subset of S

- null event - empty set
- **elementary** event - singleton

Probability Axioms

28 October 2017 18:07

set function P:
- takes a set as argument
- returns value

for set function to be a probability,

for any event $E \subseteq S$:
- $0 \leq P(E) \leq 1$
- $P(S) = 1$
- if $E \cap F = \emptyset$ (disjoint / mutually exclusive) then
  - $P(E \cup \mathrm{F}) = P(E) + P(F)$
- Otherwise, if not disjoint:
  - $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
    - addition rule - takes away overlap

For any two events E and F

- $\max\big(P(E), P(F)\big) \leq \mathrm{P(E \cup F)} \leq \mathrm{P(E) + P(F)}$
- $P(E) + P \ \ - 1 \leq P(E \cap F) \leq \min(P(E), P(F))$

Conditional Probability

28 October 2017 18:08

P(A/B)

probability A occurs given that B has occurred

- $P(A|B)=\frac{P(A\cap B)}{P(B)}$
  - gives intersection as a % of area of B

- $P(\bar{A}|B) = 1 - P(A|B)$

Multiplication Law:

- $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$

Independence:

- $P(A|B) = P(A)P(B)$

Probability Tables

28 October 2017 18:08

Separate rows or columns are disjoint

obtain **conditional probability** from rows or columns as well

Total Probability

28 October 2017 18:08

Form an unconditional probability from conditional ones

A_i , i=0…j , form a partition of S, then,

- $P(B) = \sum_{i=1}^{k} P\left(B|A_i\right) P(A_i)$

Bayes Theorem

28 October 2017 18:08

<u>Switching conditional probabilities:</u>

- $P(A_k|B) = \dfrac{P(B|A_k)P(A_k)}{P(B)}$

  - probability that A_k occurred if B occurred

Can represent P(B) as a total probability

- $P(A_k|B) = \dfrac{P(B|A_k)P(A_k)}{P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)}$

Discrete Random Variables

28 October 2017 17:16

Maps a Sample space S, into R - a set of real numbers

Theoretical Mean - Expected Value:

- $E(X) = \sum_x x\, f_X(x) = \sum_x x\, P(X = x)$
- population mean, $\mu = E(X)$
- Properties:
  - $E(aX + b) = \boldsymbol{a}E(X) + \boldsymbol{b}$
  - $E(X + Y) = E(X) + E(Y)$

Probability Mass Function - returns probability of P(X=x) - noted by $f_X(x)$

CMF - $F_X(x)$ Cumulative version of PMF, used for P(X < x_k)

PMF can vary by another variable - noted by $\boldsymbol{f_X(x;\theta)}$

Theoretical Variance:

- $Var[X] = E\left[(X-\mu)^2\right] = E\left(X^2\right) - E(X)^2 \geq 0$
- population variance, σ^2
- Properties:
  - $Var(aX + b) = \boldsymbol{a^2}Var(X)$
  - $Var(X \pm Y) = Var(X) + Var(Y)$
    - for uncorrelated X,Y

Uniform Distribution

29 October 2017 13:20

- PMF: $f_X(x) = \frac{\mathbf{1}}{\boldsymbol{k}}$, for $x = 1, 2, \dots, k$

- CDF: $F_X(x) = \frac{\boldsymbol{x}}{\boldsymbol{k}}$

- $E(X) = \frac{\boldsymbol{k+1}}{\mathbf{2}}$
  - more generally, $E(X) = \frac{\boldsymbol{a+b}}{\mathbf{2}}\ for\ x: a \dots b$

- $Var[X] = \frac{k^2 - 1}{12}$

Binomial Distribution

29 October 2017 13:54

<u>Requires:</u>

- two **mutually exclusive** outcomes
- n **identical, independent** trials
- **constant probability** of success, p

$\underline{X \sim Bin(n, p)}$

- random variable X is binomially distributed
- **n** number of trials
- **p** probability of success

- $f_X(x; p, n) = \boldsymbol{\binom{n}{x} p^x (1-p)^{n-x}}$
- $E(X) = \boldsymbol{np}$
- $Var[X] = \boldsymbol{np(1-p)}$

Geometric Distribution

29 October 2017 14:16

REVISE A LEVEL GEO SERIES PROPERTIES

'how many trials before the first success'

$\underline{X \sim Geo(p)}$

X - counts number of trials until first success
p - probability of a success

- $f_X(x;p) = \boldsymbol{(1-p)^{x-1}p}$, $\quad for\ x = 1,2,\ldots$
- $E(X) = \boldsymbol{\frac{1}{p}}$
- $Var[X] = \boldsymbol{\frac{1-p}{p^2}}$

Poisson Distribution

29 October 2017 14:26

Requirements

- independent events in time and space

$\underline{X \sim Poisson(\lambda)}$

- $\lambda > 0$
- $f_X(x; \lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, $for\ x = \mathbf{0}, 1, 2, \ldots$
- $E(X) = Var(X) = \lambda$

Continuous Random Variables

29 October 2017 16:29

<u>Probability Density Function</u> - - noted by $f_X(x)$

- must be **non-negative**
- total **area under graph = 1**

CDF - $F_X(x)$ Cumulative version of PDF, used for P(X < x_k)

$$F_X(x_0) = P(X \leq X_O) = \int_{-\infty}^{x_O} f_X(x)dx$$

- $F_X(-\infty) = 0$
- $F_X(\infty) = 1$

$$PDF = \frac{d}{dx}CDF \quad \rightarrow \quad f_X(x) = \frac{dF_X(x)}{dx}$$

PDF can <u>vary by another variable</u> - noted by $\boldsymbol{f_X(x;\theta)}$

<u>Theoretical Mean - Expected Value:</u>

- $E(g(X)) = \int_{-\infty}^{\infty} \boldsymbol{g(x)} * \boldsymbol{f_X(x)ds}$

- Properties:
  - $E(aX+b) = \boldsymbol{a}E(X) + \boldsymbol{b}$

<u>Theoretical Variance:</u>

- $Var[X] = E[(X-\mu)^2] = E(X^2) - E(X)^2 \geq 0$

- Properties:
  - $Var(aX+b) = \boldsymbol{a^2}Var(X)$

Continuous Uniform Distribution

05 November 2017 19:38

$\underline{X \sim Unif(a,b)}$

$$f_X(x) = \begin{cases} \boldsymbol{\frac{1}{b-a}} & a \le x \le b \\ 0 & otherwisse \end{cases}$$

$$\boldsymbol{E(X) = \frac{b+a}{2}}$$

$$\boldsymbol{Var(X) = \frac{(b-a)^2}{12}}$$

$$F_X(u) = \begin{cases} 0 & u < a \\ \boxed{\boldsymbol{\frac{u-a}{b-a}}} & a \le u \le b \\ 1 & u > b \end{cases}$$

Exponential Distribution

05 November 2017 19:46

$\underline{X \sim expo(\lambda)}$

$$\boldsymbol{E(X)} = \frac{\mathbf{1}}{\boldsymbol{\lambda}}$$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & otherwise \end{cases}$$

$$\boldsymbol{Var(X)} = \frac{\mathbf{1}}{\boldsymbol{\lambda}^{\mathbf{2}}}$$

$$F_X(u; \lambda) = \begin{cases} 0 & u \leq 0 \\ 1 - e^{-\lambda u} & u > 0 \end{cases}$$

Memoryless Property:

$$P(T > b + a \mid T > a) = P(\boldsymbol{T} > \boldsymbol{b})$$

- **doesn't matter** that a has occurred

* Change of Variable

05 November 2017 20:12

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

Normal Distribution

05 November 2017 20:12

$$\boldsymbol{X \sim N(\mu, \sigma^2)}$$

$$\boxed{\boldsymbol{X = \sigma Z + \mu}}$$

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

$$\boldsymbol{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

$$Z \sim N(0,1)$$
$$\boldsymbol{\mu = 0}$$
$$\boldsymbol{\sigma^2 = 1}$$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$\Phi(z) = \int_{-\infty}^{z} \phi(u) du$$

Sampled - Var - Var / n^2

* Chebyshev's Inequality

14 January 2018 22:20

X is a random variable with

$\mu = E(X)$

$\sigma^2 = Var(X)$

then, for all $a > 0$

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

* Time-to-failure

28 October 2017 18:06

Prob of component failure =Θ

> prob component doesn't fail = $1 - \Theta$

Prob system **with n** components **functions**
In Series: = $product\ of\ each\ working$
In parallel: = $1\ - product\ of\ each\ failing$

failure time distribution $F_T(t)$
failure time density $f_T(t)$

reliability time distribution $R_T(t)$ = 1- $F_T(t)$

hazard rate = $\frac{f_T(t)}{R_T(t)}$

cumulative hazard function $H\ integral\ hazard\ rate$

Weibull

$$F_T(t;\lambda,\beta) = \begin{cases} 1 - e^{-(\lambda t)^\beta} & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Mean time to failure

MTTF = $\int_0^\infty R_T(t)dt$

Marginals

28 October 2017 18:06

independent if:

- joint PMF/PDF = product of **marginal** PMFs/PDFs

* Conditional

19 May 2018 00:44

Conditional

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Variance

$$\mathrm{Var}[X|Y=y] = \begin{cases} \sum_i [x_i - \mathrm{E}[X|Y=y]]^2 P(X=x_i|Y=y), & \text{disc} \\ \int_{-\infty}^{+\infty} [x - \mathrm{E}[X|Y=y]]^2 f_{X|Y}(x|y)dx, & \text{cont.} \end{cases}$$

For random variables X and Y,

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}[X|Y]] + \mathrm{Var}[\mathrm{E}[X|Y]]$$

Expectation

GET MARGINAL FIRST

$$= \int_{-\infty}^{\infty} x f_X(x)dx$$

Screen clipping taken: 25/05/2018 01:31

Conditional Expectation

$$g(y) = \mathrm{E}[X|Y=y] = \begin{cases} \sum_i x_i P(X=x_i|Y=y), & \text{discrete,} \\ \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx, & \text{continuous.} \end{cases}$$

Note that $\mathrm{E}(X)$ is a number while $\mathrm{E}[X|Y]$ is a random variable!

* Covariance

19 May 2018 01:03

Cov(X,Y)
> 0 - large X with large Y
<0 - large X with small Y

if independent -> cov() = 0

but cov = 0 !-> independence

Correlation

19 May 2018 01:50

$$\rho = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}$$

FROM -1 to +1

Joint Normal

19 May 2018 08:35

* Moments

19 May 2018 09:42

$$m_r = E[X^r]$$

MGF

The MGF of a random variable X is defined as

$$m_X(t) = \mathrm{E}(e^{tX}) = \begin{cases} \sum_x e^{tx} f_X(x) & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx & X \text{ continuous} \end{cases}$$

Two independent RVs

$$\begin{aligned} m_{X+Y}(t) &= \mathrm{E}(e^{t(X+Y)}) = \mathrm{E}(e^{tX} e^{tY}) \\ &= \mathrm{E}(e^{tX})\, \mathrm{E}(e^{tY}) = m_X(t) m_Y(t)\,. \end{aligned}$$

Power Series

$$\begin{aligned} m_X(t) = \mathrm{E}(e^{tX}) = \mathrm{E}\left(\sum_{r=0}^{\infty} \frac{(tX)^r}{r!}\right) &= \sum_{r=0}^{\infty} \frac{\mathrm{E}(X^r)}{r!} t^r \\ &= 1 + \frac{\mathrm{E}(X)}{1!} t + \frac{\mathrm{E}(X^2)}{2!} t^2 + \frac{\mathrm{E}(X^3)}{3!} t^3 + \ldots \\ &= 1 + \frac{m_1}{1!} t + \frac{m_2}{2!} t^2 + \frac{m_3}{3!} t^3 + \ldots \end{aligned}$$

Differentiate

evaluate the derivatives of $m_X(t)$ at $t = 0$. First, notice that

$$\begin{aligned} \frac{d}{dt} m_X(t) &= m_1 + m_2 t + \frac{m_3}{2} t^2 + \ldots \\ \frac{d^2}{dt^2} m_X(t) &= m_2 + m_3 t + \frac{m_4}{2} t^2 + \ldots \\ &\vdots \end{aligned}$$

evaluate the derivatives of $m_X(t)$ at $t = 0$. First, notice that

$$\begin{aligned}
\frac{d}{dt}m_X(t) &= m_1 + m_2 t + \frac{m_3}{2}t^2 + \dots \\
\frac{d^2}{dt^2}m_X(t) &= m_2 + m_3 t + \frac{m_4}{2}t^2 + \dots \\
&\vdots \\
\frac{d^r}{dt^r}m_X(t) &= m_r + m_{r+1} t + \frac{m_{r+2}}{2}t^2 + \dots
\end{aligned}$$

* Two independent RV

19 May 2018 10:21

Summary: For two independent random variables X and Y,

$$
\begin{aligned}
f_{X,Y}(x,y) &= f_X(x) f_Y(y) \\
\mathrm{E}(XY) &= \mathrm{E}(X)\,\mathrm{E}(Y) \\
\mathrm{Cov}(X,Y) &= 0 \\
\rho &= 0 \\
\mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) \\
\mathrm{Var}(X-Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) \\
m_{X+Y}(t) &= m_X(t) m_Y(t) \\
\mathrm{E}[X|Y] &= \mathrm{E}(X)
\end{aligned}
$$

* Change of Variables

19 May 2018 10:26

$$Z = g(X, Y)$$

THEN

PDF of Z

$$\boldsymbol{f_Z(z) = \int_{x=-\infty}^{+\infty} f_{XY}(x, y(z, x))dx}$$

IF XY Independent

$$\boldsymbol{f_Z(z) = \int_{y=-\infty}^{+\infty} f_X(z-y) f_Y(y) dy}$$

CONVOLUTION

$$U = R(X, Y) \qquad V = S(X, Y)$$

THEN

if $X = L(U, V), Y = T(U, V)$

Jacobian

$$J = \begin{bmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial U} & \frac{\partial L}{\partial V} \\ \frac{\partial T}{\partial U} & \frac{\partial T}{\partial V} \end{bmatrix}$$

$$f_{UV}(u, v) = |\det(J)| f_{XY}(x, y)$$

Rayleigh Distribution

19 May 2018 11:52

Assume X and Y independent and both distributed $\sim N(0, \sigma^2)$, such that

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}.$$

We are interested in $U = \sqrt{X^2 + Y^2}$.
Let us add the RV $V = \tan^{-1}\left(\frac{Y}{X}\right)$ such that

$$X = U \cos V, \ \ Y = U \sin V$$

If X and Y are zero mean independent Normal random variables with common variance, then $\sqrt{X^2 + Y^2}$ has a Rayleigh distribution and $\tan^{-1}\left(\frac{Y}{X}\right)$ has a uniform distribution. Moreover these two derived RVs are independent.

Alternatively, for X and Y zero mean independent Normal random variables, $X + jY$ represents a complex Normal RV. It follows that the magnitude and phase of a complex Normal RV are independent with Rayleigh and uniform distributions respectively.

Properties

19 May 2018 13:07

- RVs $X_1, \ldots, X_n$ are independent
- RVs have the same expectation: $\mathrm{E}(X_i) = \mu, \forall i$
- RVs have the same variance: $\mathrm{Var}(X_i) = \sigma^2, \forall i$

Results focus on

$$S_n = X_1 + \ldots + X_n, \;\; \mathrm{E}\, S_n = n\mu, \;\; \mathrm{Var}\, S_n = n\sigma^2$$

$$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}, \;\; \mathrm{E}\, \bar{X}_n = \mu, \;\; \mathrm{Var}\, \bar{X}_n = \frac{\sigma^2}{n}$$

Weak Law

28 October 2017 18:07

$$U_n \xrightarrow{p} \theta.$$

$$\forall \epsilon > 0 \quad \lim_{n \to \infty} P(|U_n - \theta| > \epsilon) = 0$$

<u>Weak Law of Large Numbers</u>

$$\forall \epsilon > 0 \quad \lim_{n \to \infty} P(\left|\bar{X}_n - \mu\right| > \epsilon) = 0$$

$$\forall \epsilon > 0 \quad \lim_{n \to \infty} P(\left|\bar{X}_n - \mu\right| \leq \epsilon) = 1.$$

CLT

19 May 2018 11:17

RVs independent and identically distributed

$$X_1 + X_2 + \ldots + X_n \to \mathrm{N}(n\mu, n\sigma^2) \, .$$

Stats and Sampling

28 October 2017 18:07

Statistic is any quantity calculated from sample data

Statistic is a random variable - has a sampling distribution

$$E[\bar{X}] = \mu$$

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

$$s = \frac{\sigma}{\sqrt{n}}$$

IF n > 30
and samples are iid.

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

Recall the formula for the sample variance:

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

Point Estimation

19 May 2018 13:20

estimate value of unknown parameters (θ) in with sample data

$f(x|\theta)$

- roman symbols - sample
- Greek symbols - population parameters

Point Estimator - best guess at unknown population parameter

Estimator Properties

19 May 2018 13:20

Aim for:

- **unbiasedness**
- **minimum variance**

in the sampling distribution

Bias:

for an ubiased point estimator $\hat{\theta}$

$E[\hat{\theta}] = \theta \quad - \text{show that it equals, if know it's unbiased}$

$\hat{\theta} - \theta$ gives the bias of the estimator

Can assess bias without knowing θ if we are confident of the parent distribution

e.g. sample mean is ALWAYS unbiased

AIM for MVU Estimator

but sometimes a little bias can give much smaller variance

So, choose **estimator with smallest Mean Squared Error - MSE**

$$MSE_\theta(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right] \quad = \quad Var(\hat{\theta}) + \left\{Bias_\theta(\hat{\theta})\right\}^2$$

Generating Estimators:

- Method of Moments
- Least Squares
- Maximum likelihood

$$\hat{\theta} = \omega^n \sum_{i=1}^{n} X_i$$

$$\boldsymbol{E(\hat{\theta})} = \omega^n \sum_{i=1}^{n} E(X_i)$$

$$\boldsymbol{Var(\hat{\theta})} = \omega^{n+1} \sum_{i=1}^{n} Var(X_i)$$

Method of Moments

19 May 2018 13:20

EQUATING SAMPLE MEAN TO POPULATION MEAN

kth moment of X is $m_k = \mathrm{E}(X^k)$

Kth Sample Moment

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k\,, \quad \text{for } k = 1, 2, \ldots$$

$\hat{\theta} = corresponding\ sample\ moment$

Maximum Likelihood Estimation

19 May 2018 13:20

Likelihood function , sample size n

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

once simplified, take log

$$\ell(\boldsymbol{\theta}) = \boldsymbol{log L}(\boldsymbol{\theta})$$

Take **first derivative and equate to 0**

- get Maximum Likelihood Estimator
    - (used to get ML estimates)

Take second derivative

- confirm it is a maximum (<0)