Chapter 1. Probility and Distribution

by Yuejian Mo

May 26, 2018

用有效的方式收集数据的问题的研究， 构成了数理统计学中两个分支， 其一叫做抽样理论，其二叫做实验设计（试验设计）。

1 Some note

The number of permuations of n distinct objectes taken r at a time is

$$_nP_r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}, r = 0, 1, 2, ..., n.$$

The number of combinations of n distinct objects taken r at a time is

$$\left(\begin{array}{c} n \\ r \end{array} \right) = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}, r = 0, 1, 2, ..., n.$$

The binomial coeffient of the term of $x^r y^{n-r}$ in the expansion of

$$(x+y)^n = \sum_{r=0}^{n} \left(\begin{array}{c} n \\ r \end{array} \right) x^r y^{n-r}$$

is $\left(\begin{array}{c} n \\ r \end{array} \right)$, where n is a positive integer and r is a non-negative less than or equal to n.

The number of ways in which a set of n distinct objects can be partitioned into k subsets with n_1 objects in the first subset, n_2 objects in the second subset,..,and n_k objects in the k-th subset is

$$\left(\begin{array}{c} n \\ n_1, ..., n_k \end{array} \right) = \frac{n!}{n_1! \cdots n_k!},$$

which is the multinomial coefficient of the term of $x_1^{n_1} \cdots x_k^{n_k}$ in the expansion of $(x_1 + \cdots + x_k)^n$, where $n_1 + \cdots + n_k = n$.

Here are some useful formulae

- $\left(\begin{array}{c} x \\ r \end{array} \right) = \left(\begin{array}{c} n-1 \\ r \end{array} \right) + \left(\begin{array}{c} n-1 \\ r-1 \end{array} \right)$

2 Sample Space

An experiment is a process of observation or measurement. The resuts obtained from an experiment are called the *outcomes* of the experiment.The set of all possible outcomes of an experiment is called the *sample space* denoted by $\mathbb{S}$. Each outcome in a sample space is called an *elements* or a sample *point*. An *event* is a subset of a sample space.

Acoording to the number of elements thery obtain, sample space can be classified into *discrete* sample and *continuous* sample space. A sample space is discrete, if the number of elements is finite or countable. A sample space is continuous, if the sample space consists of a continum.

Events has operation as complement, union and intersection.

3 Properties of probability

Defintion 1.1 (Probability of a set). Let $\mathbb{A}$ be a subset of the sample space $\mathbb{S}$, then $\Pr(\mathbb{A})$ is said to be the probability of $\mathbb{A}$ if

i. $\Pr(\mathbb{A}) \geqslant 0$ and $\Pr(\mathbb{S}) = 1$;

ii. If $\mathbb{A}_1, \mathbb{A}_2, \ldots$ is a sequence of mutually exclusive of $\mathbb{S}$, then

$$\Pr\left(\bigcup_{i=1}^{\infty} \mathbb{A}_i\right) = \sum_{i=1}^{\infty} \Pr(\mathbb{A}_i)$$

4 Conditional Probability

Definition (Conditional probability of two sets). If $\mathbb{A}$ and $\mathbb{B}$ are two events in the sample space $\mathbb{S}$, the conditional probability of $\mathbb{B}$ given $\mathbb{A}$ is defined by

$$\Pr(\mathbb{B}|\mathbb{A}) = \frac{\Pr(\mathbb{A} \cap \mathbb{B})}{\Pr(\mathbb{A})}$$

where $\Pr(\mathbb{A}) > 0$.

Definiation (Independncy of two events). Two events $\mathbb{A}$ and $\mathbb{B}$ are said to be *indepentdent*, denoted by $\mathbb{A} \perp \mathbb{B}$, if

$$\Pr(\mathbb{A} \cap \mathbb{B}) = \Pr(\mathbb{A}) \times \Pr(\mathbb{B})$$

Theorem (Independcy). Let $\mathbb{A} \perp B$, then $\mathbb{A} \perp \mathbb{B}'$ and $\mathbb{A}' \perp \mathbb{B}'$.

Definition (Mutual independency). Event $\mathbb{A}_1, \ldots, A_n$ are said mutuallly independent, if the probability of the intersection of any $2, 3, \ldots,$ or n of these events equals the product of their respective probalilities.

Definition (Partition). A partition of the sample space $\mathbb{S}$ is a collection of mutually exclusive sets $\mathbb{B}_1, \ldots, B_n$, such that $\mathbb{S} = \cup_{i=1}^{n} \mathbb{B}_i$.

Bayes formula. Let $\mathbb{B}_1, \ldots, \mathbb{B}_n$ be a partition of sample space $\mathbb{S}$ and $\mathbb{A}$ be an evnet, then

i. Law of total probability:

$$\Pr(\mathbb{A}) = \sum_{i=1}^{n} \Pr(\mathbb{A}|\mathbb{B}) \times \Pr(\mathbb{B})$$

ii. Bayes formula:

$$\Pr(\mathbb{B}_j|\mathbb{A}) = \frac{\Pr(\mathbb{A}|\mathbb{B}_j) \times \Pr(\mathbb{B}_j)}{\sum_{i=1}^{n} \Pr(\mathbb{A}|\mathbb{B}_j) \times \Pr(\mathbb{B}_j)} \qquad \text{for } j = 1, \ldots, n.$$

5 Probility Distribution

Defniation (Random variable)s. A *random variable* (r.v.) is a funtion from a sample space $\mathbb{S}$ into the real numbers. An r.v. is *discrete* if it takes values in a finite or countable set. An r.v. is *continus* if it take values over some interval.(Just like we trans angle to radian)

Defination (Probability mass function). If X is a discrete r.v., the function denoted by

$$p(x) = \Pr(X = r)$$

for each x within the range of X is called the *probability mass function* (pmf) of X.

Definition (Probability density function). Let X be a continues r.v.. A non-negative function $f(x)$ is called the *probability density function*(pdf) of X, if

$$\Pr(\mathbb{A}) = \int_{\mathbb{A}} f(x)dx$$

In other word:

$$\Pr(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

Defination(Cumulative density function). The *cumulative distribution function*(cdf) of an r.v. X is defined by

$$F(x) = \Pr(X \leqslant x) = \begin{cases} \sum_{t \leqslant x} p(t) & \text{, if } X \text{ is discete,} \\ \int_{-\infty}^{x} f(x)dx & \text{, if } X \text{ is continuous.} \end{cases}$$

6 Bivariate Distributions

Definition (Bivariate pmf). If X and Y are two discrete r.v.'s, the function defined by

$$p(x, y) = \Pr(X = x, Y = y)$$

for each pair of values (x, y) within the range of X and Y is called the joint pmf of X and Y.

Similary, a bivariate function $f(x, y)$ is called a joint pdf of the continuous r.v.'s X and Y if

$$\Pr\{(X, Y) \in \mathbb{A}\} = \int\int_{\mathbb{A}} f(x, y)dxdy$$

for a region $\mathbb{A}$ in the domain of (X, Y).

Then the joint distribution(or joint cdf) of r.v.'s (X, Y) is defined by

$$\begin{aligned} F(x, y) &= \Pr(X \leqslant x, Y \leqslant y) \\ &= \begin{cases} \sum_{s \leqslant x, t \leqslant y} p(s, t) & \text{, if } X \text{ and } Y \text{ are discete,} \\ \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)dxdy & \text{, if } X \text{ and } Y \text{ are continuous.} \end{cases} \end{aligned}$$

6.1 Marginal and conditional distributions

Let $p(x, y)$ be the joint pmf of discrete r.v.'s (X, Y). The *marginal* pmfs of X and Y are defined by

$$p(x) = \sum_y p(x, y) \quad \text{and} \quad p(y) = \sum_x p(x, y),$$

respectively. The *conditional* pmfs of X given $Y=y$ and Y given $X=x$ are defined by

$$p(x|y)=\frac{p(x,y)}{p(y)}, p(y)\neq 0 \quad \text{and} \quad p(y|x)=\frac{p(x,y)}{p(x)}, p(x)\neq 0$$

respectively.

6.2 Independency of two random variables

Let $f(x,y)$ denote the joint pdf of r.v's (X,Y), and $f(x)$ and $f(y)$ be their marginal pdfs. The r.v.'s X and Y are said to be *independent*, denoted by $X \perp Y$, if

$$\begin{array}{rcll} f(x,y) &=& f(x)\times f(y) & ,\forall (x,y)\in \mathcal{S}_{(X,Y)}, \quad \text{or} \\ F(x,y) &=& F(x)\times F(y) & ,\forall (x,y)\in \mathcal{S}_{(X,Y)}. \end{array}$$

where $\mathcal{S}_{(X,Y)} \hat{=} \{(x,y)\colon f(x,y)>0\}$ denotes the joint *support* of (X,Y).

6.3 Expecation, Variance and Moments

The expectation of $g(X)$ is defined by

$$E\{g(X)\}=\begin{cases} \sum_x g(x)p(x) & \text{, if } X \text{ is discrete} \\ \int_{-\infty}^{+\infty} f(x)\, g(x)dx & \text{, if } X \text{ is continuous.} \end{cases}$$

When $g(X)=X$, the expectation of X, measure the *central location* of the pdf of X. (how about expectation of multi-r.v.) Let $\mu=E(X)$, then

$$\sigma^2=\text{Var}(X)=E(X-\mu)^2=E(X^2)-\mu^2$$

is a measure of the *dispersion* of the pdf of X. $\sigma=\sqrt{\text{Var}(X)}$ is called the standard deviation. We also define covaiance as

$$\text{Cov}(X_1,X_2)=E\{(X_1-\mu_1)(X_2-\mu_2\}, \text{where } \mu_i=E(X_i), i=1,2.$$

Covariance also can be calculated by $\text{Cov}(X_1,X_2)=E(X_1X_2)-E(X_1)E(X_2)$.

The r-th moment of the r.v. X is defined by $\mu_r'=E(X^r)$. The r-th central moment of the r.v. X is defined by $\mu_r=E[(X-\mu)^r]$. $\mu_3=E[(X-\mu)^3]$ is a measure of *asymmertry* of the pdf of X. The fourth central moment $\mu_4=E(X-\mu)^4$ is a measure of kurtosis(峰态), which is the *degree* of flatness of a density near its center.

7 Moment Generating Function

For an r.v. X, if $E(e^{tX})$ exists for nay $t\in(-\varepsilon,\varepsilon)$ for some $\varepsilon>0$, then

$$M_X(t)=E(e^{tX})$$

is called teh *moment generating function*(mgf) of X. By using Maclaurin's expansion, we have

$$M_X(t) = E\left\{\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right\} = \sum_{n=0}^{\infty} E\left(\frac{t^n}{n!}X^n\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!}E(X^n)$$

Thus, the n-th moment can be obtained by

$$\mu_n' = E(X^n) = \frac{d^n M_X(t)}{dt^n}\bigg|_{t=0}.$$

8 Useful Distribution

8.1 Bivariate Normal Distribution

It is well know that X is normally disributed with mean μ and variance σ^2, denoted by $X \sim N(\mu, \sigma^2)$, if its pdf is

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, -\infty < x < +\infty$$

To introduce the bivariate normal distribution, first of all, we define the *correlation coefficient* of X_1 and X_2 by

$$\rho = \mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1 \sigma_2}.$$

A random vector $\mathbf{x} = (X_1, \ldots, X_d)^T$ is said to follow a d-dimemsional normal distribution, if its joint pdf if given by

$$N_d(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

for $\boldsymbol{x} \in \mathbb{R}^d$, where the mean vecotr $\boldsymbol{\mu} \in \mathbb{R}^d$ and the covariance matrix $\boldsymbol{\Sigma}$ is postive definite, denoted by $\boldsymbol{\Sigma} > 0$. We wil write $\mathbf{x} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

8.2 Beta distribution and Gamma distribution

The gamma function

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

which is well defined for $\alpha > 0$.

$$X \sim \mathrm{Beta}(x|a, b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}, 0 \leqslant x \leqslant 1.$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. $E(X) = \frac{a}{a+b}$. The beta distribution is the conjugate prior for the binomial likehood.

$$X \sim \mathrm{Gamma}(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x \geqslant 0.$$

$E(X) = \frac{\alpha}{\beta}$,

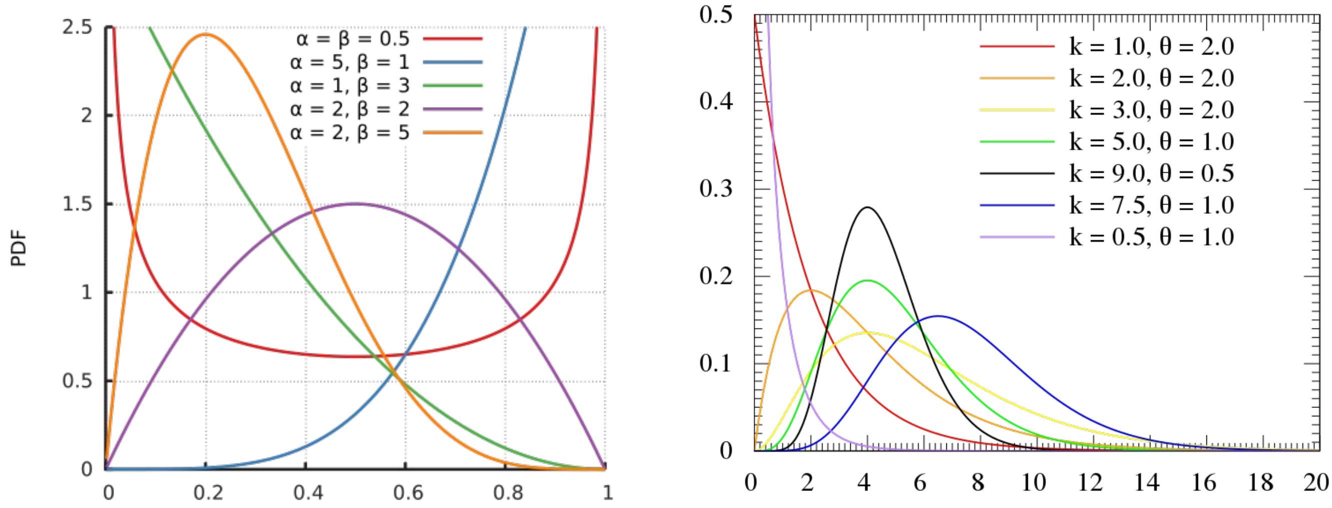

Figure 1. The beta(left) and gamma(right) distribution with different parameter.

8.3 Categorical Distribution

8.4 Zero-inflated Possion Distribution

9 Inverse Byes Formulae

1. *Point-wise formula*

$$\begin{aligned} & f_Y(y) = \frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)} f_X(x) \\ \Rightarrow \; & \int_{\mathcal{S}_Y} f_Y(y)dy = \int_{\mathcal{S}_Y} \frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)} f_X(x)dy = f_X(x) \int_{\mathcal{S}_Y} \frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)} dy = 1 \\ \Rightarrow \; & f_X(x) = \left\{ \int_{\mathcal{S}_Y} \frac{f_{(Y|X)}(y|x)}{f_{(X|Y|)}(x|y)} dy \right\}^{-1}, \text{for any } x \in \mathcal{S}_X \end{aligned}$$

2. *Function-wise formula*

   Substituting point-wise formula to $f_Y(y) = \frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)} f_X(x)$, we can get(symmetry)

$$f_X(x) = \left\{ \int_{\mathcal{S}_X} \frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x)} dx \right\}^{-1} \frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x|)}$$

   for all $x \in \mathcal{S}_X$ and an arbitrarily fixed $y_0 \in \mathcal{S}_Y$.

3. *Sampling-wise formula*

   By dropping the normalizing constant in functione-wise formula, we obtain

$$f_X(x) \propto \frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x)}$$

   For discrete r.v.,the point-wise formula and smaple-wise formula for all $x \in \mathcal{S}_Y$and fiexed $y_0 \in \mathcal{S}_Y$.

$$\Pr(X = x) = \left\{ \sum_{y \in \mathcal{S}_Y} \frac{\Pr(Y = y|X = x)}{\Pr(X = x|Y = y)} \right\}^{-1}, \Pr(X = x) \propto \frac{\Pr(X = x|Y = y_0)}{\Pr(Y = y_0|X = x)}$$

10 References

- https://en.wikipedia.org/wiki/Central_moment
- https://en.wikipedia.org/wiki/Beta_distribution