Chapter 2.1 Discrete Distribution（离散分布）

Section 2.1 **Random variable** of the discrete type

Outcomes of experiment
- Numerical
  e.g. Rolling a 6-sided die
- Not numerical
  e.g. Flipping a coin

For the latter case, we can define a function X to associate the outcomes with numerical values.

Example1: Rolling a die: S={1, 2, 3, 4, 5, 6}

X(i)=i, i=1,2,3,4,5,6.

Flipping a coin: S={H, T}

X(H)=0, X(T)=0.5

def def

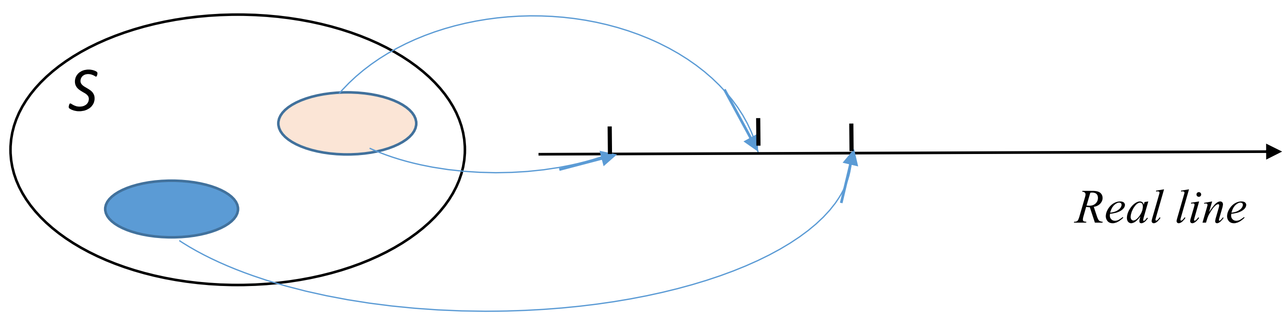
Definition 2.1-1 [Random Variable (RV)]

Given a random experiment with Sample Space S, a function X:
S→B ⊆ R that assign one and only one real number $X(s)=x$ to each s∈S is called **random variable**.
In other words, A ***random variable*** is a function from a sample space S into the real numbers.

Definition [discrete random variable] →离散型随机变量

The range of X is the set B={ x| $X(s)=x$, s∈S}. A RV is called **discrete** if its range B is finite or countable.

$$X: \ S \to B \subseteq \mathbb{R}$$



Given a experiment with Sample space

$$S=\{s_1, \dots, s_n\}$$

with a probability function P on S and we define a random variable X with range B $=\{x_1, \dots, x_m\}$. we can define a probability function P on B in the following way:

$$P(X = x_i) \triangleq P(\{X = x_i\}) = P_r(\{s_j \,|\, X(s_j) = x_i,\ \ s_j \in S\})$$

$$P(x_i \in A) \triangleq P(\{x_i \in A\}) = P_r(\{s_j \,|\, X(s_j) \in A,\ \ s_j \in S\})$$

Note that $A \subset B$

A note on notation:

Random variables will always be denoted with uppercase letters And the numerical values of RV will be denoted by the corresponding lowercase letters

$e.g.\quad X \to RV \qquad\qquad e.g.\quad x \to numerical\ \ value\ \ of\ \ RV.$

Thus, the random variable X can take the value x

Definition 2.1-2 [probability mass function (pmf)]

Suppose that X: S→B ⊆ R is a discrete random variable. Then a function f(x): B →[0,1] is called pmf, if

(a) $f(x)>0$, $x \in B$

(b) $\sum_{x \in B} f(x) = 1$

(c) $P(x \in A)=\sum_{x \in A} f(x)$ where $A \subset B$

- We often extend the definition domain of $f(x)$ from B to R and let $f(x)=0$ for $x \notin B$.
- B is the **range** of X and is also called the **support** of $f(x)$.
- From now on, we consider pmf $f(x)$: R →[0,1].

Definition [Cumulative distribution function (cdf)]

The *cumulative distribution function* or *cdf* of a random variable X, denoted by $F(x)$, is defined by

$$F(x) = P(X \le x) \triangleq P(\{s \mid X(s) \le x, \;\; s \in S\})$$

$$x \in (-\infty, +\infty), \;\; cdf \;\; is \;\; often \;\; called \;\; the \;\; distribution \;\; function \;\; of \;\; X$$

Definition [uniform distribution]
When a pmf is constant over the support.

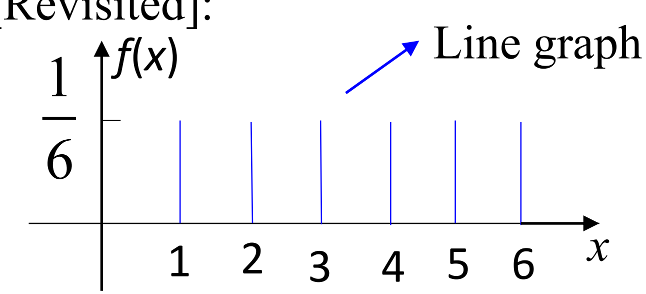
Example 2: Rolling a die:

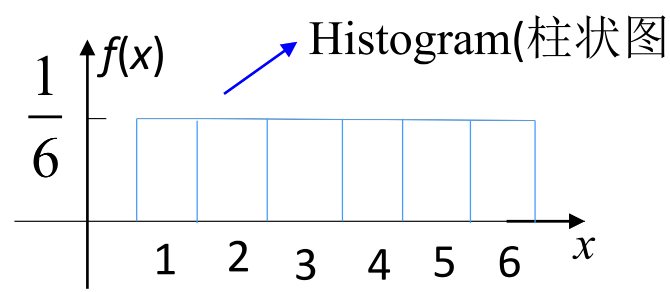
S={A, B, C, D, E, F}
→B={1,2,3,4,5,6}

- $define\ \ a\ \ RV\ \ X(s)=x\ \ for\ \ \forall s\in S$
- $pmf\ \ f(x)=\begin{cases}1/6, & if\ \ x\in B\\ 0, & if\ \ x\notin B\end{cases}$
- $cdf\ \ F(x)=P(X\le x)=\begin{cases}0, & if\ \ x<1\\ k/6, & if\ \ k\le x<k+1,\ \ k=1,2,3,4,5\\ 1, & if\ \ x\ge 6.\end{cases}$

Definition [line graph]
A line graph of the pmf $f(x)$ of the random variable X is a graph having a vertical line segment drawn from $(x,0)$ to $[x, f(x)]$ at each $x\in$ B.

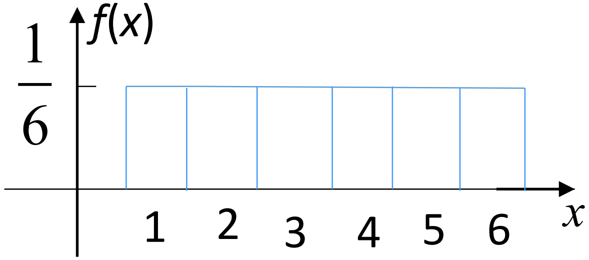
Example2 [Revisited]:





Definition [probability histogram]

If X assume only integer values, a probability histogram of pmf $f(x)$ is a graphical representation that has a rectangle of height $f(x)$ and a base of length 1, centered at x for each $x \in S$.



Probability Histogram

Section 2.2 Mathematical exception

We will learn many probability distributions, it's important to introduce concepts in summarizing their key characteristics.

$$\begin{cases} expectation \\ variance \\ \cdots \end{cases}$$

Analogy would be of great use.

- Motivation example. (Page 57)

A man proposes a game: let the other player throw a die and the player receives payment as follows:

$$A = \{1, \ 2, \ 3\} \rightarrow 1 \ dollar$$
$$B = \{4, \ 5\} \rightarrow 2 \ dollars$$
$$C = \{6\} \rightarrow 3 \ dollars$$

$$X: \ S = A \cup B \cup C \ \rightarrow \ X(S) = \{1, \ 2, \ 3\}$$

Now let X be a RV to represent the payment, the pmf of X is:

$$f(x) = \frac{4-x}{6}, \quad x = 1, 2, 3$$
$$f: \ X(S) \rightarrow [0,1]$$

The man charge the player 2 dollars for each play. Can the man make profit if the game is repeated endlessly?

$$\text{Solution: Payment of } \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ occur } \begin{bmatrix} 3/6 \\ 2/6 \\ 1/6 \end{bmatrix} \text{ of the times.}$$

$$\text{The average payment is } \boxed{1\times 3/6 + 2\times 2/6 + 3\times 1/6 = 5/3.} \triangleq \boldsymbol{E(X)}$$

Longrun average value of X

$$\text{So the man can earn } 2 - 5/3 = 1/3 \text{ per play on average.}$$

More generally, we are interested in long run average value of a function of X, say g(X)

Definition [Expectation]

Assume X is a discrete RV with range space X(S) and $f(x)$ is its pmf. If $\sum_{x\in X(S)} g(x)f(x)$ exists, then it's called the **expectation** or the **expected value** of g(X) and is denoted by E[g(X)]. That is,

$$E[g(X)] = \sum_{x\in X(S)} g(x)f(x)$$

Example 1 (Page 59):

$$Let\ X\ be\ a\ RV\ with\ X(S)=\{-1,\ 0,\ 1\}\ and\ its\ pmf\ is\ f(x)=\frac{1}{3}$$

$$for\ \forall x\in X(S).\ What's\ E(x^2)?$$

$$Solution:\ E(x^2)=\sum_{x\in X(S)} x^2 f(x)=(-1)^2\times 1/3+0^2\times 1/3+1^2\times 1/3=2/3.$$

- Properties of mathematical expectation:

Theorem 2.2-1 **[Page 60]**

$$Consider\ a\ RV\ X:S\to X(S),\ and\ its\ pmf\ f:\ X(S)\to[0,1].$$

When the mathematical expectation exists, it satisfies the following properties:

(a) If c is a constant, $E(c)=c$

(b) If c is a constant, and $g(x)$ is a function,

$$E[cg(x)]=cE[g(x)].$$

(c) If c_1 and c_2 are constants, $g_1(x)$ and $g_2(x)$ are functions,

$$E[c_1g_1(x)+c_2g_2(x)]=c_1E[g_1(x)]+c_2E[g_2(x)]$$

Mathematical expectation is a linear operator.

Example 2 (Page 61):

$Let\ \ g(x)=(x-b)^2\ \ where\ \ b\ \ is\ \ a\ \ constant\ \ to\ \ be\ \ chosen.$

$and\ \ suppose\ \ E\left[(X-b)^2\right]\ \ exists.\ Find\ \ the\ \ value\ \ of\ \ b\ for\ \ which$

$$E\left[(X-b)^2\right]\ \ is\ \ minimal$$

$Solution:$

$$E\left[(X-b)^2\right]=E\left[X^2-2bX+b^2\right]$$

$$=E(X^2)-2bE(X)+b^2\triangleq h(b)$$

$$\frac{\partial h(b)}{\partial b}=-2E(X)+2b;\qquad \frac{\partial^2 h(b)}{\partial b^2}=2>0.$$

$$So\ \ when\ \ \frac{\partial h(b)}{\partial b}=0,\ \ E\left[(X-b)^2\right]\ \ is\ \ minimal.$$

$\Rightarrow b=E(X).$ → Mean is the MMSE(误差平方和均值最小) estimator.

Section 2.3 Special mathematical exception

- **Mean of RV:** The expectation of X is also called the mean of X.

$$E(X) = \sum_{x \in X(S)} x f(x) \overset{X(S)=\{u_1,\dots,u_k\}}{=} \sum_{i=1}^{k} u_i f(u_i)$$

Mechanic Interpretation:

$u_i \to The\ distance\ of\ ith\ point\ from\ the\ origin.$

$f(u_i) \to The\ weight\ of\ the\ ith\ point.$

$u_i f(u_i) \to A\ moment\ having\ a\ moment\ arm\ of\ length\ u_i$

$E(X) \to The\ first\ moment\ about\ the\ system;\ the\ \boxed{centroid.}$ → 质心

$Why\ E(X)\ is\ the\ centroid?$

$If\ we\ choose\ E(X)\ as\ the\ new\ origin,\ then\ we\ compute\ the\ first\ moment\ agaon:$

$$E\left[X - E(X)\right] = E(X) - E(X) \qquad \leftarrow because\ E(X)\ is\ constant$$

$$= 0.$$

$Hence\ The\ first\ moment\ about\ the\ E(X)\ is\ zero. \Rightarrow E(X)\ is\ centroid.$

- Variance of RV:

$$Var(X) = E\left[(X - E(X))^2\right] = \sum_{x \in X(S)} (x - E(X))^2 f(x) = E\left[X^2 - 2XE(X) + E^2(X)\right]$$

$$= E(X^2) - 2E(XE(X)) + E^2(X)$$

$$= E(X^2) - 2E(X) \cdot E(X) + E^2(X)$$

$$= E(X^2) - \left[E(X)\right]^2$$

The positive square root of the variance is called the **standard deviation** (δ)

Example 1 (Page 66):

Let X equal to the number of spots after a $6-$*sided die is rolled.*

A reasonable probability model is:

$$f(x) = P(X = i) = 1/6, \qquad i = 1, 2, 3, 4, 5, 6.$$

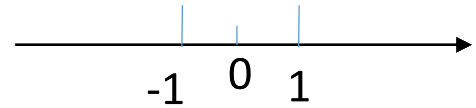
Mean of X: $E(X) = 1/6 \times (1 + 2 + 3 + 4 + 5 + 6) = 7/2.$

Variance of X: $Var(X) = E(X - E(X))^2 = E(X^2) - \left[E(X)\right]^2 = \dfrac{91}{6} - \dfrac{49}{4} = \dfrac{35}{12}.$

Example 2 [Interpretational standard deviation] (Page 66):

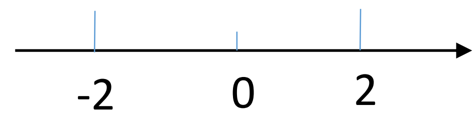
$$X \quad has \quad pmf \quad f(x) = 1/3 \quad for \quad x = -1, 0, 1$$

$$E(X) = 0; \quad Var(X) = 2/3; \quad \delta_X = \sqrt{2/3}.$$



$$Y \quad has \quad pmd \quad f(y) = 1/3 \quad for \quad y = -2, 0, 2$$

$$E(Y) = 0; \quad Var(Y) = 8/3; \quad \delta_Y = 2\sqrt{2/3}.$$



Standard deviation is a measure of the dispersion or spread of the points belonging to the range space of RV.

- Properties of variance:

Let X be a RV

$$Var(c) = 0; \qquad Var(cX) = c^2 Var(X).$$

➢ rth moment of the distribution

- $Let\ r\ be\ a\ positive\ integer.\ If\ E(X^r) = \sum_{x \in X(S)} x^r f(x)\ exists\ (finite),\ then$

$it\ is\ called\ the\ rth\ moment\ of\ the\ distribution\ about\ the\ origin.$

- $In\ addition,\ E\left[(X-b)^r\right] = \sum_{x \in X(S)} (x-b)^r f(x)$

$is\ called\ the\ rth\ moment\ of\ the\ distribution\ about\ b.$

- $E\left[(X)_r\right] \triangleq E\left[X(X-1)(X-2)\cdots(X-r+1)\right] \quad \to \quad rth\ factorial\ moment$

- Mean
- Variance
- Standard deviation
- moments

→ Characteristics of distribution of probability

We now define a function that will help us generate the moments of a distribution:

Definition [Moment generating function (mgf)]

Let X be a discrete RV with range space X(S). If there exists h>0 such that $E(e^{tX}) = \sum_{x \in X(S)} e^{tx} f(x) \quad exists\ and\ is\ finite\ for\ -h < t < h$

Then the function defined by $M(t) = E(e^{tX})$ is called the **moment-generating function** of X

- Properties of mgf:
  - *I.* $M(0)=1$
  - II. If two RVs have the same mgf, they must have the same distribution of probability.

Example 3:

If X has the mgf

$$M(t) = e^t(\frac{3}{6}) + e^{2t}(\frac{2}{6}) + e^{3t}(\frac{1}{6}), \quad -\infty < t < +\infty$$

then the support of the pmf $f(x)$ *of* X *is* $X(S)=\{1, 2, 3\}$ *and the associated pmf* $f(x) = \frac{4-x}{6}, \ x = 1,2,3.$

$$M'(t) = \sum_{x\in X(S)} xe^{tx} f(x)$$

$$M''(t) = \sum_{x\in X(S)} x^2 e^{tx} f(x)$$

$$M^{(r)}(t) = \sum_{x\in X(S)} x^r e^{tx} f(x)$$

Noted question:

- Is $M(t)$ differentiable? in 1st, 2nd,…order
- Interchange of the differentiation and summation.

Putting $t=0$ *we find* $M'(0)=E(X)$; $M''(0)=E(X^2)$; $M^{(r)}(0)=E(X^r)$

Observation: The moments can be computed by differentiating $M(t)$!

Example 4 [Page 71]:

Suppose X has the geometric distribution, that is, the pmf of X is

$$f(x) = q^{x-1}p \qquad x = 1,2,3,\ldots,n,\ldots \qquad p = 1-q.$$

Then the mgf of X is

$$M(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx}\cdot q^{x-1}\cdot p = \frac{p}{q}\sum_{x=1}^{\infty}\left(qe^{t}\right)^{x}$$

$$= \frac{p}{q}\left[(qe^{t}) + (qe^{t})^{2} + \cdots\right]$$

$$= \frac{p}{q}\frac{qe^{t}}{1-qe^{t}} = \frac{pe^{t}}{1-qe^{t}} \qquad \textit{provided} \quad qe^{t} < 1 \Leftrightarrow t < \ln q.$$

Let h=-lnq that is positive. To find the mean and variance of X,

$$M'(t) = \frac{pe^{t}}{1-qe^{t}} - \frac{(pe^{t})(-qe^{t})}{(1-qe^{t})^{2}} = \frac{pe^{t}}{(1-qe^{t})^{2}}$$

$$M''(t) = \frac{pe^{t}(1+qe^{t})}{(1-qe^{t})^{3}}$$

⇒

$$M'(0) = E(X) = \frac{p}{(1-q)^{2}} = \frac{1}{p}$$

$$M''(0) = E(X^{2}) = \frac{1+q}{p^{2}}$$

$$Var(X) = E(X^{2}) - \left[E(X)\right]^{2} = \frac{1+q}{p^{2}} - \frac{1}{p^{2}} = \frac{q}{p^{2}}$$