MTH113TC: Intro. to Probability and Statistics

Lesson 3 - Random variables and their distributions - Part I

minerva

Dr. Yinna Ye

Department of Mathematics, XJTLU

Suzhou, PR China

1. Introduction

2. Discrete random variables

1. Introduction

Introduction (1)

When an experiment is performed, we are interested mainly in some function of the outcome as opposed to the actual outcome itself.

Example: A fair coin is flipped twice: $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$. For $\omega \in \Omega$, let $X(\omega)$ be the number of heads, so that

$$X((H,H)) = 2, \quad X((H,T)) = X((T,H)) = 1, \quad X((T,T)) = 0.$$

Introduction (2)

The function $X = X(\omega)$, $\omega \in \Omega$, is a random variable which represents the number of heads in this experiment.
Because the value of the random variable X is determined by the outcomes of the experiment, we may assign probabilities to the possible values of X:

$$\mathbb{P}(X = 0) = \mathbb{P}(\{(T, T)\}) = \frac{1}{4},$$

$$\mathbb{P}(X = 1) = \mathbb{P}(\{(H, T), (T, H)\}) = \frac{1}{2},$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\{(H, H)\}) = \frac{1}{4}.$$

Since X must take on one of the values 0, 1, 2, we must have

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{i=0}^{2} \{X = i\}\right) = \sum_{i=0}^{2} \mathbb{P}(X = i).$$

Introduction (3)

Example: Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total number of n flips is made. If we let $X = X(\omega)$, $\omega \in \Omega$, denote the number of times the coin is flipped, then X is a random variable taking on one of the values $1, 2, 3, \ldots, n$ with respective probabilities

Introduction (4)

$$\mathbb{P}(X=1)=\mathbb{P}(\{(H)\})=p,$$
$$\mathbb{P}(X=2)=\mathbb{P}(\{(T,H)\})=(1-p)p,$$
$$\mathbb{P}(X=3)=\mathbb{P}(\{(T,T,H)\})=(1-p)^2p,$$
$$\vdots$$
$$\mathbb{P}(X=n-1)=\mathbb{P}(\{(\underbrace{T,T,\dots,T,}_{(n-2)\text{ times}}H)\})=(1-p)^{n-2}p,$$
$$\begin{aligned}\mathbb{P}(X=n)=&\mathbb{P}(\{(\underbrace{T,T,\dots,T,}_{(n-1)\text{ times}}T)\}\text{ or }\{(\underbrace{T,T,\dots,T,}_{(n-1)\text{ times}}H)\})\\=&(1-p)^n+(1-p)^{n-1}p\\=&(1-p)^{n-1}(1-p+p)\\=&(1-p)^{n-1}\end{aligned}$$

Introduction (5)

As a check, note that

$$
\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n}\{X=i\}\right) &= \sum_{i=1}^{n}\mathbb{P}(X=i) \\
&= \sum_{i=1}^{n-1} p(1-p)^{i-1} + (1-p)^{n-1} \\
&= p\left[\frac{1-(1-p)^{n-1}}{1-(1-p)}\right] + (1-p)^{n-1} \\
&= 1-(1-p)^{n-1} + (1-p)^{n-1} \\
&= 1.
\end{aligned}
$$

There are two important class of random variables, namely **discrete random variables** and **continuous random variables**.

2. Discrete random variables

Discrete random variables (1)

Definition

A function $X : \Omega \to M$ is called a **random variable**. If the codomain M is a finite or countable set, then X is called a **discrete random variable**.

Remark

The random variable X's defined in the examples of the previous section are all discrete random variables.

Discrete random variables (2)

Definition (Probability mass function)

Let $\mathbb{P}$ be a probability defined on the sample space Ω, and $X : \Omega \to M$ be a discrete random variable. The function $f_X : \mathbb{R} \to [0, 1]$ defined by

$$f_X(x) := \begin{cases} \mathbb{P}(X = x) & , \quad \text{if } x \in M, \\ 0 & , \quad \text{if } x \notin M; \end{cases}$$

is called the **probability mass function (p.m.f) of the random variable** X.

Discrete random variables (3)

Example (Revisit Example in Page 4)

In a coin twice flips experiment, let X denotes number of heads. All the possible values that X can take is in $M = \{0, 1, 2\}$. And so the p.m.f of X is

$$f_X(x) = \begin{cases} \frac{1}{4}, & \text{if } x = 0, 2 \\ \frac{1}{2}, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Discrete random variables (4)

Example (Revisit Example in Page 6)

In the experiment described in Page 6, all the possible values that X can take is in $M = \{1, 2, \ldots, n\}$. And so the p.m.f. of X is

$$f_X(k) = \begin{cases} (1-p)^{k-1}p, & \text{for } k = 1, 2 \ldots, n-1 \\ (1-p)^{k-1}, & \text{for } k = n \\ 0, & \text{otherwise} \end{cases}$$

Theorem (Properties of probability mass functions)

Let $X : \Omega \to M$ be a discrete random variable with probability mass function f_X. Then the following hold:

(1) *Positivity:* $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$.

(2) *Normalization:* $\sum_{x \in M} f_X(x) = 1$.

Discrete random variables (5)

Proof.

(1) Since $0 \leq \mathbb{P}(X = x) \leq 1$, clearly $f_X(x)$ is nonnegative.

(2) Suppose $M := \{x_1, x_2, \cdots, x_m\}$, where $m = |M|$. Note that the events $\{X = x_j\}$ form a partition of Ω. Hence

$$
\begin{aligned}
&\sum_{x \in M} f_X(x) \\
\overset{\text{def.}}{=} &\sum_{x \in M} \mathbb{P}(X = x) \\
= &\,\mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + \cdots + \mathbb{P}(X = x_m) \\
\overset{\text{Ax. 3}}{=} &\,\mathbb{P}\left(\{X = x_1\} \cup \{X = x_2\} \cup \ldots \cup \{X = x_m\}\right) \\
= &\,\mathbb{P}(\Omega) \overset{\text{Ax. 1}}{=} 1.
\end{aligned}
$$

$\square$

Discrete random variables (6)

Definition (Cumulative distribution function)

Let $X : \Omega \to M$ be a discrete random variable with probability mass function f_X. Then the function $F_X : \mathbb{R} \to [0, 1]$, defined by

$$F_X(x) := \mathbb{P}(X \leq x)$$

is called **cumulative distribution function (c.d.f) of the random variable X.**

Remark

The relationship between f_X and F_X of a discrete random variable X is

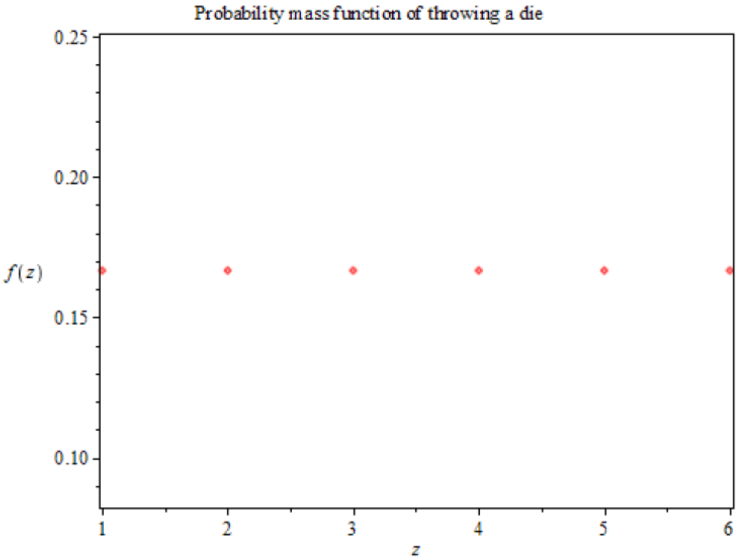
$$F_X(x) = \sum_{\substack{y \in M \\ y \leq x}} f_X(y).$$

Discrete random variables (7)

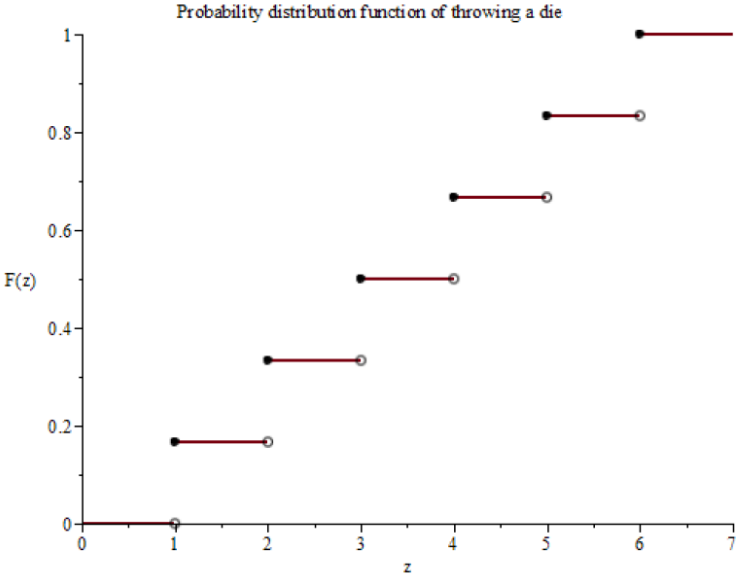
Example: The p.m.f of rolling a fair die with Z denoting the face number is

$$f_Z(z) := \begin{cases} 1/6, & \text{if } z = 1, \\ 1/6, & \text{if } z = 2, \\ 1/6, & \text{if } z = 3, \\ 1/6, & \text{if } z = 4, \\ 1/6, & \text{if } z = 5, \\ 1/6, & \text{if } z = 6, \\ 0, & \text{otherwise.} \end{cases}$$

Discrete random variables (8)



Discrete random variables (8)



Discrete random variables (9)

The c.d.f $F_Z(z) = \mathbb{P}(Z \leq z)$ of rolling a fair die with Z denoting the face number is

$$F_Z(z) := \begin{cases} 0, & \text{if } z < 1, \\ f_Z(1) = 1/6, & \text{if } 1 \leq z < 2, \\ f_Z(1) + f_Z(2) = 1/3, & \text{if } 2 \leq z < 3, \\ f_Z(1) + f_Z(2) + f_Z(3) = 1/2, & \text{if } 3 \leq z < 4, \\ f_Z(1) + f_Z(2) + f_Z(3) + f_Z(4) = 2/3, & \text{if } 4 \leq z < 5, \\ f_Z(1) + f_Z(2) + f_Z(3) + f_Z(4) + f_Z(5) = 5/6, & \text{if } 5 \leq z < 6, \\ f_Z(1) + f_Z(2) + f_Z(3) + f_Z(4) + f_Z(5) + f_Z(6) = 1, & \text{if } z \geq 6. \end{cases}$$

2.1. Statistic characteristic of discrete random variables

Definition (Expectation/Mean)

Let X be a discrete random variable with values in M and probability mass function f_X . The number

$$\mathbb{E}[X] := \sum_{x \in M} x f_X(x) = \sum_{x \in M} x \mathbb{P}(X = x) \tag{1}$$

is called the **expectation (or expected value, mean)** of the random variable X.

Example (Revisit Example in Page 16)

Note that $M = \{1, 2, 3, 4, 5, 6\}$

$$\begin{aligned}\mathbb{E}(Z) &= \sum_{z \in M} z \cdot f_Z(z) \\ &= 1 \cdot f_Z(1) + 2 \cdot f_Z(2) + \cdots + 6 \cdot f_Z(6) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{7}{2}.\end{aligned}$$

Example

We say that I is an indicator variable for the event A if

$$I = \begin{cases} 1, & \text{if } A \text{ occurs;} \\ 0, & \text{if } A^c \text{ occurs.} \end{cases}$$

Find $\mathbb{E}(I)$.

Solution: Since $f_I(1) = \mathbb{P}(I = 1) = \mathbb{P}(A)$ and $f_I(0) = \mathbb{P}(I = 0) = \mathbb{P}(A^c)$, we have

$$\mathbb{E}[I] = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A).$$

That is the expectation of the indicator variable for the event A is equal to probability that A occurs.

Remark:

(1) When each value x has equal probability, i.e., $f_X(x) = \frac{1}{|M|}, \forall x \in M$, then by using the Fundamental Formula the expectation reduces to the common **arithmetic mean**:

$$\mathbb{E}[X] := \sum_{x \in M} x f_X(x) = \frac{1}{|M|} \sum_{x \in M} x.$$

(2) Note that the expectation of a random variable may be finite or infinite. If M is a finite subset of $\mathbb{R}$, the sum at the right most side of (1) is finite and in this case we call the mean exists. If M is countably infinite, then the series maybe divergent and thus the mean becomes infinite and we say it does not exist.

Statistic characteristic of discrete random variables (5)

(3) More generally, given a discrete random variable X and a function $h : M \to \mathbb{R}$. The **expectation of** $h(X)$ is defined as

$$\mathbb{E}[h(X)] := \sum_{x \in M} h(x) f_X(x).$$

(4) For instance, if we choose $h(X) = aX + b$, where $a, b \in \mathbb{R}$ are constants. Then we have

$$\begin{aligned}
\mathbb{E}(aX + b) &\overset{\text{def.}}{=} \sum_{x \in M} (ax + b) f_X(x) \\
&= a \sum_{x \in M} x f_X(x) + b \underbrace{\sum_{x \in M} f_X(x)}_{=1 \text{ by Normal.}} \\
&= a\mathbb{E}[X] + b \cdot 1 = a\mathbb{E}(X) + b.
\end{aligned}$$

Or more generally, we have the **linear property** of taking expectations, i.e.,

$$\mathbb{E}[ah(X) + bg(Y)] = a\mathbb{E}[h(X)] + b\mathbb{E}[g(Y)],$$

where h, g are two given functions and X, Y are two discrete random variables.

Definition (Variance and Standard deviation)

Let X be a discrete random variable with values in M and probability mass function f_X . The number

$$\mathsf{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in M} (x - \mathbb{E}[X])^2 f_X(x) \quad (2)$$

is called the **variance of** X, subject to the RHS of (2) is finite. Since the variance is non-negative, $\sqrt{\mathsf{Var}(X)}$ exists and is called the **standard deviation of** X.

Remark

(1) The variance and standard deviation measure the spread of the random variable about its mean.

(2) An alternative formula for Var(X) is derived as follows: let $\mu = \mathbb{E}(X)$,

$$\begin{aligned}
\text{Var(X)} &\overset{\text{def.}}{=} \mathbb{E}[(X-\mu)^2] \\
&= \sum_{x\in M}(x-\mu)^2 f_X(x) \\
&= \sum_{x\in M}(x^2 - 2\mu x + \mu^2) f_X(x) \\
&= \underbrace{\sum_{x\in M} x^2 f_X(x)}_{=\mathbb{E}[X^2]} - 2\mu \underbrace{\sum_{x\in M} x f_X(x)}_{=\mu} + \mu^2 \underbrace{\sum_{x\in M} f_X(x)}_{=1 \text{ by Normal.}} \\
&= \mathbb{E}[X^2] - \mu^2.
\end{aligned}$$

That is,

$$\mathrm{Var(X)} = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{3}$$

Example (Revisit Example in Page 16)

Calculate $\mathrm{Var}(Z)$ if Z represents the outcome when a fair die is rolled.

<u>Solution</u>: It is shown in Page 22 that $\mathbb{E}[Z] = \frac{7}{2}$. Also,

$$\begin{aligned}\mathbb{E}[Z^2] &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} \\ &= \frac{91}{6}.\end{aligned}$$

Hence by (3),

$$\mathrm{Var}(Z) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Theorem

Let X and Y be two discrete random variables.

(1) *Then for any constants $a, b \in \mathbb{R}$,*

$$\mathit{Var}(aX + b) = a^2 \mathit{Var}(X).$$

(2) *Two random variables X and Y are said to be independent if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*
If X and Y are independent, then for any constants $a, b \in \mathbb{R}$,

$$\mathit{Var}(aX + bY) = a^2 \mathit{Var}(X) + b^2 \mathit{Var}(Y).$$

Proof:

(1) By definition of variance, we have

$$\begin{aligned}
\mathrm{Var}(aX+b) = & \sum_{x\in M}(ax+b-\mathbb{E}[aX+b])^2\cdot f_X(x) \\
\overset{\text{Linearity}}{=} & \sum_{x\in M}(ax+b-a\mathbb{E}(X)-b)^2\cdot f_X(x) \\
= & \ a^2\sum_{x\in M}(x-\mathbb{E}[X])^2\cdot f_X(x) \\
= & \ a^2\mathrm{Var}(X).
\end{aligned}$$

(2) The proof is left as exercise.

2.2. Specific types of discrete random variables

2.2.1. Bernoulli and Binomial random variables

Definition (Bernoulli random variables)

A random variable X take values 1 and 0 with probabilities $p \in (0, 1)$ and q $(:= 1 - p)$, respectively, is called a **Bernoulli random variable** with parameter p. Sometimes we think of these values as representing the "success rate" or the "failure rate" of a trial. The p.m.f. is

$$f_X(0) = 1 - p, \quad f_X(1) = p.$$

Experiments involving a Bernoulli variable are called **Bernoulli trials**.

Example (Bernoulli trials)

(1) Tossing a fair coin with $p = 1/2$.

(2) A medical treatment maybe effective with a probability p and ineffective with probability $(1-p)$.

Theorem

For a Bernoulli random variable X with parameter p,

$$\mathbb{E}[X] = p, \quad \mathit{Var}(X) = p(1-p).$$

Bernoulli and Binomial random variables (3)

Proof.

We have

$$\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

and

$$\mathbb{E}(X^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p,$$

from which we have

$$\mathsf{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = p - p^2 = p(1-p).$$

$\square$

Definition

When perform n independent Bernoulli trials, i.e. X_1, X_2, $\ldots$, X_n are independent and identically distributed (i.i.d) Bernoulli random variables with parameter p and count the total number of successes $Y := X_1 + X_2 + \cdots + X_n$. Then Y is called a **Binomial random variable with parameters** (n, p), where $n \in \mathbb{N}$ and $p \in (0, 1)$. Equivalently, Y is called a Binomial random variable with parameters (n, p), if the p.m.f of Y is given by

$$f_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{4}$$

The experiment generating a binomial variable is called a **binomial experiment**.

Remark

(1) The definition of Binomial random variable (4) may be verified by first noting that the probability of the particular sequence of n outcomes containing k success and $(n-k)$ failures is $p^k(1-p)^{n-k}$ by the assumed independence of trials. Equation (4) then follows, since there are $\begin{pmatrix} n \\ k \end{pmatrix}$ different sequence of the n outcomes leading to k successes and $(n-k)$ failures. This can most easily be seen by noting that there are $\begin{pmatrix} n \\ k \end{pmatrix}$ different choices of the k trials that result in sucess.

(2) If Y is a binomial random variable with parameters (n, p), then we use the shorthand notation:

$$Y \sim B(n, p).$$

(3) The normalization property of the p.m.f f_Y given by (4) can be verified as follows:

$$\sum_{k=0}^{n} f_Y(k) = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k} \overset{\text{Bino. formu.}}{=} [p+(1-p)]^n = 1.$$

Theorem

Let Y be a binomial random variable with parameters (n, p). Then we have

$$\mathbb{E}[Y] = np, \quad \mathit{Var}(Y) = np(1-p).$$

Proof.

By linear property of expectation, we have

$$\begin{aligned}\mathbb{E}[Y] =& \mathbb{E}[X_1 + X_2 + \cdots + X_n] \\ =& \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X_1] = np.\end{aligned}$$

By the theorem in Page 29, we obtain

$$\begin{aligned}\mathsf{Var}(Y) =& \mathsf{Var}(X_1 + X_2 + \cdots + X_n) \\ =& 1^2 \cdot \mathsf{Var}(X_1) + 1^2 \cdot \mathsf{Var}(X_2) + \cdots + 1^2 \cdot \mathsf{Var}(X_n) \\ =& n \cdot \mathsf{Var}(X_1) = np(1-p).\end{aligned}$$

$\square$

2.2.2. Geometric distribution

Geometric distribution (1)

Definition

A **Geometric distribution with parameter** $p \in (0, 1)$ is a random varibale with the geometric mass function:

$$f(k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots \tag{5}$$

Geometric distribution (2)

Remark

Geometric distribution arises in the following way. Suppose that independent Bernoulli trials (parameter p) are performed at times $1, 2, \ldots$ Let W be the time which elapses until the first success. Then the event

$\{W = k\} = \{$At the kth trial, success occurs for the first time$\}$.

So by the the independence of the trials,

$$f(k) = \mathbb{P}(W = k) = (1-p)^{k-1}p.$$

Geometric distribution (3)

Example

An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability that

(a) exactly n draws are needed?

(b) at least k draws are needed?

Geometric distribution (4)

Solution: If we let X denote the number of draws needed to select a black ball, then X satisfies Equation (5) with $p = M/(M+N)$. Hence,

(a)

$$\mathbb{P}(X=n) = \left(\frac{N}{M+N}\right)^{n-1} \frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}.$$

(b)

$$\mathbb{P}(X \geq k) = \sum_{n=k}^{\infty} \mathbb{P}(X=n) = \frac{M}{M+N} \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1}$$

$$= \frac{\frac{M}{M+N}\left(\frac{N}{M+N}\right)^{k-1}}{1-\frac{N}{M+N}} = \left(\frac{N}{M+N}\right)^{k-1}.$$

Geometric distribution (5)

Exercise

Prove that if X follows Geometric distribution with parameter p, then

$$\mathbb{E}[X] = \frac{1}{p}, \quad \mathsf{Var}(X) = \frac{1-p}{p^2}.$$

2.2.3. Poisson random variable

Poisson random variable (1)

Definition

A random variable X that takes one of the values $0, 1, 2, \ldots$ is said to be a **Poisson random variable with parameter** λ, if for some $\lambda > 0$ the p.m.f. is given by

$$f_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2\ldots \tag{6}$$

Remark:

(1) The equation (6) defines a probability mass function, since $f_X(k) \geq 0$, for $k = 0, 1, \ldots$ and

$$\sum_{k=0}^{\infty} f_X(k) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

(2) If X is a Poisson random variable with parameter λ, then we use the shorthand notation $X \sim Poi(\lambda)$.

Poisson random variable (2)

(3) Poisson distribution focuses on the number of discrete events or occurrences over a specified interval or continuum (time, length, distance, etc. For instance, see the following example).

Example

Suppose that the number of typographical errors on a single page of this paper has a Poisson distribution with parameter $\lambda = 1/2$. Calculate the probability that there is at least one error on this page.

Solution: Letting X denote the number of errors on this page, we have

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-1/2} \approx 0.393.$$

Poisson random variable (3)

Theorem

If $X \sim Poi(\lambda)$, then

$$\mathbb{E}(X) = \lambda, \quad Var(X) = \lambda.$$

Proof:

The mean can be obtained through the following calculation:

$$\begin{aligned}
\mathbb{E}[X] &:= \sum_{i=0}^{\infty} i\mathbb{P}(X = i) = \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} = e^{-\lambda} \sum_{i=1}^{\infty} \frac{i\lambda^i}{i!} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}, \quad \text{by letting } j = i - 1 \\
&= \lambda, \quad \text{since } \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}.
\end{aligned}$$

Poisson random variable (4)

To determine its variance, we first compute $\mathbb{E}[X^2]$.

$$
\begin{aligned}
\mathbb{E}[X^2] :=& \sum_{i=0}^{\infty} i^2 \mathbb{P}(X = i) = \sum_{i=0}^{\infty} \frac{i^2 e^{-\lambda} \lambda^i}{i!} \\
=& \sum_{i=1}^{\infty} \frac{i^2 e^{-\lambda} \lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} \frac{i e^{-\lambda} \lambda^{i-1}}{(i-1)!} \\
=& \lambda \sum_{j=0}^{\infty} \frac{(j+1) e^{-\lambda} \lambda^j}{j!}, \quad \text{by letting } j = i - 1 \\
=& \lambda \left[\underbrace{\sum_{j=0}^{\infty} \frac{j e^{-\lambda} \lambda^j}{j!}}_{=\lambda} + \underbrace{\sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!}}_{=1} \right] = \lambda(\lambda + 1),
\end{aligned}
$$

Thus, we obtain $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda$.

The end of Lesson 3 - Part I