Probability

1. Review.

• experiment, trial:
the process of obtaining an observed result of some phenomenon.
• outcome : observed result

<u>Definition</u>. *The set of all possible outcomes of an experiment is called the sample space, denoted by S.*

<u>Definition</u>. *If a sample space S is either finite or countably, then it is called a discrete sample space. Otherwise, is called a continuous sample space.*

<u>Definition</u>. *An event is a subset of the sample space S. If A is an event, then "A occurred" if "A contains the out come that occurred"*

<u>Definition</u>. *For a given experiment, S denotes the sample space and $A_1, \cdots$ represent possible events. A set function that associates a real value $P(A)$ with each event A is called a probability set function, and $P(A)$ is called the probability of A, if the following properties are satisfied:*
i) $0 \leq p(A)$ *for every* A
ii) $P(S) = 1$
iii) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ *if* $A_1, \cdots$ *are pairwise mutually exclusive events.*

Example.
Throwing a coin, the outcome is "head" or "tail", the sample space is $\{head, tail\}$, we can give it an event A like "a head exists", the $A = \{head\} \subseteq S$

2. Discrete Random Variable.

Definition. *A random variable, say X, is a function defined over a sample space S, that associated a real number with each possible outcome in S*

$$X(e) = x, \text{ where } e \in S$$

Definition. *A random variable that can take on at most a countable number of possible values is said to be discrete. For a discrete random variable X, we define the probability mass function $p(a)$ of X by*

$$p(a) = P\{X = a\}$$

Property. $f(x_i) \geq 0, \; \sum_{\text{all } x_i} f(x_i) = 1$

Definition. *If X is a discrete random variable having a probability mass function $p(x)$, then the expectation, or the expected value, of X, denoted by $E[X]$, is defined by*

$$E(X) = \sum_x x f(x)$$

and we some times denote $E(X) = \mu$

Note. Don't just remember that the expected value just the mean.

Example. If we define $X(\text{"}head\text{"}) = 20$, $X(\text{"}tail\text{"}) = 100$, we get $E[X] = 60$ but it cannot give us more information.

Theorem. *If X is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then, for any real-valued function g,*

$$E[g(X)] = \sum_i g(x_i) p(x_i)$$

Proof. By grouping together all the terms in $\sum_i g(x_i)p(x_i)$

$$\begin{aligned}
\sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\
&= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\
&= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\
&= \sum_j y_j P\{g(X)=y_j\} \\
&= E[g(X)]
\end{aligned}$$

∎

Corollary. *If a and b are constants, then*

$$E[aX+b] = aE[X]+b$$

Definition. *If X is a random variable with mean μ, then the variance of X, denoted by $Var(X)$, is defined by*

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

Prop. An alternative formula for $\mathrm{Var}(X)$ is derived as follows:

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

Proof. think $(X-\mu)^2$ as $g(X)$, use the theorm above you get

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X-\mu)^2] \\
&= \sum_x (x-\mu)^2 p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\
&= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}$$

∎

3. Bernoulli and Binomial Random Variables.

Definition. *A trial whose outcome can be classified as either a success or a failure is performed. If we let $X = 1$ when the outcome is a success and $X = 0$ when it is a failure, then the probability mass function of X is given by*

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p \\ p(1) &= P\{X = 1\} = p \end{aligned}$$

and the random variable X is said to be a Bernoulli random variable.

Definition. *Suppose now that n independent trials, each of which results in a success with probability p or in a failure with probability $1-p$, are to be performed. If X represents the* ***number of successes*** *that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p).*

The probability mass function of a binomial random variable having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i} \mid i = 0, 1, \cdots, n$$

Note. A Bernoulli random variable is just a binomial random variable with parameters $(1, p)$.

Example. We have 10 bulbs in the cases, we know that 2 of 10 is broken. If we pick the bulbs randomly (put it back after pick), what is the probability distribution?

Solution.

x	0	1	2	3
p(x)	$C_0^3 \left(\frac{2}{10}\right)^0 \left(\frac{8}{10}\right)^3$	$C_1^3 \left(\frac{2}{10}\right)^1 \left(\frac{8}{10}\right)^2$	$C_2^3 \left(\frac{2}{10}\right)^2 \left(\frac{8}{10}\right)^1$	$C_3^3 \left(\frac{2}{10}\right)^3 \left(\frac{8}{10}\right)^0$

$\blacksquare$

The Expect Value & Variance of Binomial

X is a binomial random variable with parameters (n, p).

$$E(X) = np \;\; \& \;\; \mathrm{Var}(x) = np(1-p)$$

Proof. Just use the definition of $E(X)$ and $\mathrm{Var}(X)$

$$\begin{aligned}
E(X) &= \sum_{i}^{n} x_i p(x_i) \\
&= \sum_{i}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i}^{n} i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} p^i (1-p)^{n-i} \\
&= n \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!(n-i)!} p^i (1-p)^{n-i} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(i-1)(n-i)!} p^{i-1} q^{n-k} \\
&= np \sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\
&= np
\end{aligned}$$

■

Example. A shooter whose shooting rate is 0.6, today he shoot 100 times, what is the expect value and variance?

Solution. $E[X] = 0.6 \times 100 = 60$, $\mathrm{Var}(X) = 100 \times (1 - 0.6) \times 0.6$ ■

4*. Moment generating function (Discrete).

Definition. *The moment generating function $M(t)$ of the random variable X is defined for all real values of t by*

$$M(t) = E\left[e^{tX}\right] = \sum_x e^{tx} p(x)$$

if X is discrete with mass function $p(x)$

Mean. We call $M(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $M(t)$ and then evaluating the result at $t = 0$. For example,

$$\begin{aligned} M'(t) &= \frac{d}{dt} E[e^{tX}] \\ &= E[\frac{d}{dt}(e^{tX})] \\ &= E[Xe^{tX}] \end{aligned}$$

where we have assumed that the interchange of the differentiation and expectation operators is legitimate. That is, we have assumed that

$$\frac{d}{dt}\left[\sum_x e^{tx} p(x)\right] = \sum_x \frac{d}{dt}\left[e^{tx} p(x)\right] dx$$

in the discrete case

In general, the nth derivative of $M(t)$ is given by

$$M^n(t) = E\left[X^n e^{tX}\right]$$

implying that

$$M^n(0) = E\left[X^n\right] \;\; n \geq 1$$

Example. If X is a binomial random variable with parameters n and p, then

$$\begin{aligned} M(t) &= E\left[e^{tX}\right] \\ &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

where the last equality follows from the binomial theorem. Differentiation yields

$$M'(t) = n(pe^t + 1 - p)^{n-1} pe^t$$

Thus,

$$E[X] = M'(0) = np$$

Differentiating a second time yields

$$M''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$

so

$$E[X^2] = M''(0) = n(n-1)p^2 + np$$

The variance of X is given by

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np(1-p) \end{aligned}$$

5. Poisson Random Variable.

<u>Definition</u>. *A random variable X that takes on one of the values $0, 1, 2, \cdots$ is said to be a Poisson random variable with parameter λ if, for some $\lambda > 0$,*

$$p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!} \mid i = 0, 1, 2, \cdots$$

and above defines a probability mass function, since

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1$$

The Expect Value & Variance of Poisson
X is a Poisson random variable with parameter λ

$$E(X) = \lambda \ \& \ \mathrm{Var}(X) = \lambda$$

Theorem. *We can think poisson is a binomial distribution with parameter (n,p) where $n \to \infty$*

Proof. To see this, suppose that X is a binomial random variable with parameters (n, p), and let $\lambda = np$.Then

$$\begin{aligned} P\{X = i\} &= \frac{n!}{(n-i)!i!}p^i(1-p)^{n-i} \\ &= \frac{n!}{(n-1)!i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \end{aligned}$$

Now, for n large and λ moderate,

- $\dfrac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1$
- $\dfrac{\lambda^i}{i!} \approx e^{-\lambda}$
- $\left(1-\dfrac{\lambda}{n}\right)^i \approx 1$

Hence, for n large and λ moderate,

$$P\{X = i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}$$

■