STAT 240
Personal Notes

Marcus Chan (UW 25)

Taught by Ying-Li Qin

Chapter 1: What is Probability?

RANDOM EXPERIMENTS (1.1)

1. A "random experiment" is the process of obtaining a random observed result.
2. Random experiments can be split into two types:
   ① Controlled experiments; and
   eg flipping a coin, rolling a die
   ② Observational studies.
   eg # of students taking STAT 240 in F2021

FEATURES OF RANDOM EXPERIMENTS

Note that random experiments have the following common features:

① The outcomes/results cannot be predicted with certainty; and
② All the possible outcomes are known beforehand with certainty.

SAMPLE SPACE (1.2)

OUTCOME

An "outcome" is an observed result of interest from a random experiment.
eg the number rolled after rolling a die.

SAMPLE SPACE

The "sample space" of a random experiment is the set of all possible distinct outcomes of said experiment.
eg when rolling a 6-sided die:
$S = \{1,2,3,4,5,6\}$

EVENTS

1. An "event" of a random experiment is a group or set of outcomes of said experiment; ie subsets of the sample space.
2. There are two types of events:
   ① Simple events – consist of one outcome
   eg rolling a 1 on a 6-sided die:
   $S = \{1, \ldots, 6\}$ $\quad E = \{1\}$
   ② Compound events – consist of multiple outcomes
   eg rolling an odd number on a 6-sided die:
   $S = \{1, \ldots, 6\}$ $\quad E = \{1,3,5\}$
3. Note that
   ① Two simple events will never occur simultaneously;
   eg can never roll a 1 & 3 at the same time with one die
   ② A compound event occurs if and only if one of "its" simple events occurs; and
   eg odd # rolled ⟺ 1 rolled or 3 rolled or 5 rolled (on a 6-sided die)
   ③ Two compound events can occur simultaneously.
   eg 3 rolled ⟹ $\begin{cases} \text{odd number rolled } (E=\{1,3,5\}); \text{ and} \\ \text{multiple of 3 rolled } (E=\{3,6\}). \end{cases}$

DEFINITIONS OF PROBABILITY (1.3)

"Probability" is a quantitative measure of how likely an event is to occur.

CLASSICAL DEFINITION

1. The "classical definition" of probability states that each distinct outcome in the sample space is equally likely to occur.
2. In this case, the probability of an event E is equal to

$$P(E) = \frac{\#\text{ of ways event can occur}}{\#\text{ of outcomes in the sample space}}.$$

eg roll a 6-sided die once.
E = number is odd.
⟹ $E = \{1,3,5\}$, $\quad S = \{1,2,3,4,5,6\}$.
So $P(E) = \frac{3}{6} = \frac{1}{2}$.

RELATIVE FREQUENCY DEFINITION

The "relative frequency" definition of probability states that the probability of an event occurring is the proportion it occurs in a very long series of repetitions of the experiment.
eg rolling a 6-sided die 300 times
⟹ 3 shows up 49 of those 300 times
⟹ so $P(\text{die} = 3) \approx \frac{49}{300} \approx 1/6$.

SUBJECTIVE PROBABILITY DEFINITION

1. In the "subjective probability" definition of probability, the probability of an event is determined by an opinion (ie what a person thinks the probability is).
   eg the probability of COVID-19 being eradicated by 2022.
2. Note that this plays a role in fields like "Bayesian Statistics".

DISCRETE PROBABILITY MODELS (1.4)

In discrete probability models:
① The sample space S satisfies $|S| \leq |\mathbb{N}|$;
ie there are either a finite or countably infinite number of basic events; and
② Each probability p_i satisfies $0 \leq p_i \leq 1$; and
③ The probabilities of each basic event sum to 1;
ie $\sum p_i = 1$.

CLASSIC DISCRETE MODELS (1.5)

In classic discrete models:
① The sample space S satisfies $|S| < |\mathbb{N}|$ (ie it is finite); and
② All basic events are equally likely to occur;
ie $P(a_1) = \cdots = P(a_{|S|}) = \frac{1}{|S|}$.

Chapter 2: Counting Techniques

FULL FACTORIAL : $n!$ (2.1)

The factorial of n, denoted as "$n!$" and defined to be

$$n! = n(n-1)\cdots 1$$

is the number of ways to put n distinguishable objects in a row.

COMBINATIONS : C_n^r OR ${}^nC_r$ (2.2)

"n choose r", denoted as "C_n^r" or "${}^nC_r$", defined to be

$$C_n^r = \frac{n!}{r!(n-r)!} = \frac{n(n-1)\cdots(n-(r-1))}{r!}$$

is the number of ways to select r objects from n distinguishable objects.

PERMUTATIONS : P_n^r OR ${}^nP_r$ (2.3)

"P_n^r" or "${}^nP_r$", defined to be

$$P_n^r = \frac{n!}{(n-r)!} = n(n-1)\cdots(n-(r-1)) = C_r^n \cdot r!$$

is the number of ways to select r objects from n distinguishable objects and put them in a row.

GENERALIZATION OF COMBINATIONS (2.4)

We can show the number of ways to arrange n objects in a row, where n_1 objects are of type 1, n_2 objects are of type 2, .. , n_k objects are of type k, where $n_1 + n_2 + \cdots + n_k = n$, is

$$\text{\# of outcomes} = \frac{n!}{n_1! \cdots n_k!} = C_n^{n_1} C_{n-n_1}^{n_2} C_{n-n_1-n_2}^{n_3} \cdots C_{n_k+n_{k-1}}^{n_{k-1}} C_{n_k}^{n_k}.$$

eg Roll a die 4 times. Find P(the sum = 10).

Sol$\underline{n}$. This is equivalent to distributing 10 balls into 4 sections, where each section has at least 1 ball.

9 different spaces for the "dividers", 4 "dividers"

$\Rightarrow$ C_4^9 ways of "positioning" the dividers.

But, we exclude the options where one of the section has 7 balls, ie 7111, 1711, 1171, 1117.

Hence $P(\text{event}) = \dfrac{C_4^9 - 4}{6^4}$, since there are 6^4 outcomes of rolling a 6 sided die twice. ☒

STARS & BARS WITHOUT "EMPTY" SECTIONS

Given n stars, the # of ways to divide them up into k sections with $k-1$ rods without one of the sections containing zero elements is

$$\# = C_{k-1}^{n-1}.$$

☆ | ☆ | ☆ ☆ | ☆

eg $n=5$, $k=4$

STARS & BARS WITH "EMPTY" SECTIONS

Given n stars, the # of ways to divide them up into k sections with $k-1$ rods with one (or more) sections containing zero elements is

$$\# = C_{k-1}^{n+k-1}.$$

☆ || ☆ ☆ | ☆ ☆

eg $n=5$, $k=4$
2nd section has no elements.

Chapter 3: Probability Rules

RELATIONS AMONGST EVENTS (1.1)

EVERY EVENT $\subseteq S$ (THE "CERTAIN" EVENT)

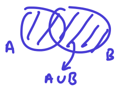
Let A be an event.
Then necessarily
$A \subseteq S = \{$ the event that always occurs $\}$.

$\emptyset$ (THE "IMPOSSIBLE" EVENT)

We use "$\emptyset$" to denote the event that never occurs.

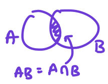
UNION OF EVENTS : $A \cup B$

Let A, B be events.
Then "$A \cup B$" is the event that at least one of the two occurs.



INTERSECTION OF EVENTS : $A \cap B$

Let A, B be events.
Then, "$A \cap B$" is the event that both A & B occur.
We also denote $A \cap B = AB$.



MUTUALLY EXCLUSIVE / DISJOINT

Let A, B be events.
Then, we say A & B are "mutually exclusive" (or "disjoint") if $A \cap B = \emptyset$.

INCLUSION : $A \subseteq B$

Let A, B be events.
Then, we say "$A \subseteq B$" if B occurs whenever A occurs; ie
A occurs $\Rightarrow$ B occurs.

COMPLEMENT : $A^c = \bar{A}$

Let A be an event.
Then, $\bar{A}$ is the event such that
$\bar{A}$ occurs $\Leftrightarrow$ A does not occur.

PARTITION OF S

Let $B_1, \dots, B_n$ be events.
Then, we say $B_1, \dots, B_n$ form a "partition" of S if
$B_1 \cup \dots \cup B_n = S$ & $B_i \cap B_j = \emptyset$ $\forall i \neq j$.

PROBABILITY RULES (1.2)

A probability function $P : \mathcal{P}(S) \to [0,1]$ is any function that satisfies the following for any $A, B \subseteq S$:

1. $P(\emptyset) = 0$;
2. $P(S) = 1$;
3. $P(A) \geq 0$ $\forall A \subseteq S$; (non-negativity)
4. $A \subseteq B \Rightarrow P(A) \leq P(B)$;
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; & (addition law of probability)
   - this generalizes to more variables as well.
6. $P(A^c) = 1 - P(A)$.

Chapter 4: Conditional Probability and Event Independence

CONDITIONAL PROBABILITY (1.1)

Let A, B be events.
Then, the probability that A happens given B already happens, denoted as "$P(A|B)$", is equal to

$$P(A|B) = \frac{P(A\cap B)}{P(B)}.$$

*note $P(B) \neq 0$ necessarily.

INDEPENDENCE OF TWO EVENTS (1.2)

1. Let A, B be events.
Then, we say A & B are "independent" if and only if

$$P(A\cap B) = P(A)\,P(B).$$

2. Note that if $P(A), P(B) \neq 0$, then A & B cannot be mutually exclusive (ie $P(A\cap B)=0$) if they are independent.

3. If A & B are independent, then
① A & B^c are independent;
② A^c & B are independent; and
③ A^c & B^c are independent.

4. Note that independence arises from independent random events.

INDEPENDENCE OF > TWO EVENTS (1.3)

1. Let $A_1, \dots, A_n$ be n events.
Then, we say $A_1, \dots, A_n$ are (mutually) independent if

$$P(A_{n_1}A_{n_2}\cdots A_{n_k}) = P(A_{n_1})\cdots P(A_{n_k}) \quad \forall \{n_1,\dots,n_k\} \in \mathcal{P}(\{1,\dots,n\}).$$

2. For the $n=3$ case, A_1, A_2 & A_3 are independent if
① $P(A_1A_2) = P(A_1)P(A_2)$;
② $P(A_1A_3) = P(A_1)P(A_3)$;
③ $P(A_2A_3) = P(A_2)P(A_3)$; and
④ $P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$.

$A_1,\dots,A_n$ ARE INDEPENDENT $\Rightarrow P(\prod_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i | A_1\cdots A_{i-1})$ (THE MULTIPLICATION FORMULA) (1.4.1)

Let $A_1, \dots, A_n$ be independent events.
Then necessarily

$$P(A_1\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1}).$$

Proof. Note that for any $k=1,\dots,n$, we have

$$P(A_k | A_1\cdots A_{k-1}) = \frac{P(A_1\cdots A_{k-1}A_k)}{P(A_1\cdots A_{k-1})} = P(A_k).$$

The proof follows trivially. ∎

$A_1,\dots,A_n$ PARTITION $S \Rightarrow P(B) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i)$ (TOTAL PROBABILITY FORMULA) (1.4.2)

Let $A_1, A_2, \dots$ form a partition of S, ie we have that $A_iA_j = \emptyset \ \forall i\neq j$ & $\bigcup_{i=1}^{\infty} A_i = S$.
Let B be an event. Then necessarily

$$P(B) = P(BS) = \sum_{i=1}^{\infty} P(A_iB) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i).$$

* this also works for finite collections of events as well.

$A_1,\dots,A_n$ PARTITION $S \Rightarrow P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{\sum_{j=1}^{\infty} P(A_j)P(B|A_j)}$ (THE BAYES FORMULA) (1.4.3)

Let $A_1, A_2, \dots$ form a partition of S, and let B be such that $P(B)\neq 0$.
Then necessarily, for any $i\in\mathbb{N}$, we have that

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{\infty} P(A_j)P(B|A_j)}.$$

* again, this also generalises to the finite case.

Chapter 5: Discrete Random Variables and Probability Models

RANDOM VARIABLES (1.1)

RANDOM VARIABLE (RV) (D1)

1. Let S be a sample space. Then, a "random variable" is defined to be some $X: S \to \mathbb{R}$.
2. Note that we usually denote random variables by capital letters. (eg X, Y, Z, etc)

DISCRETE [r.v]

1. Let $X \in \mathbb{R}^S$ be a r.v. Then, we say X is "discrete" if $|\text{range}(X)| \leq |\mathbb{N}|$.

PROBABILITY MASS FUNCTION (PMF)

1. Let $X \in \mathbb{R}^S$ be a r.v. Then, the "probability mass function" (or p.m.f) of X is defined to be the function $f: \text{range}(X) \to [0,1]$ by $f(x) = P[X=x]$ $\forall x \in \text{range}(X)$.
2. By construction of f, note that $\sum_{x \in X} f(x) = 1$.

CUMULATIVE DISTRIBUTION FUNCTION (CDF)

1. Let $X \in \mathbb{R}^S$ be a r.v. Then, the "cumulative distribution function" (or "cdf") of X is defined to be the function $F: \mathbb{R} \to [0,1]$ by $F(x) = P[X \leq x]$ $\forall x \in \mathbb{R}$.
2. Properties of cdf:
   ① $F(x_1) \leq F(x_2) \iff x_1 \leq x_2$ $\forall x_1, x_2 \in \mathbb{R}$; and
   ② $\lim_{x \to -\infty} F(x) = 0$ & $\lim_{x \to +\infty} F(x) = 1$.

PMF CAN BE OBTAINED BY CDF, AND VICE VERSA

1. Let $X \in \mathbb{R}^S$ be discrete. Then, given the pmf f of X, we can obtain X's cdf F, and vice versa.

Proof. Let $x \in \text{range}(X)$. See that
$f(x) = P[X=x] = P[X \leq x] - P[X \leq x-\varepsilon] = F(x) - F(x-\varepsilon)$,
where $\varepsilon > 0$ is such that $\text{range}(X) \cap [x-\varepsilon, x] = \{x\}$.
(Since X is discrete, such an ε will exist.)

FINDING PMF (E1)

1. Let X be the number of heads after flipping a fair coin n times. Find the pmf of X.

Sol^{n}. See that $\text{range}(X) = \{0, 1, \dots, n\}$.
Then
$$P[X=k] = C_n^k \left(\tfrac{1}{2}\right)^k \left(\tfrac{1}{2}\right)^{n-k} = C_n^k \left(\tfrac{1}{2}\right)^n$$
and so the pmf of X is $f: \{0, \dots, n\} \to [0,1]$ given by
$$f(k) = P[X=k] = C_n^k \left(\tfrac{1}{2}\right)^n \quad \forall k = 0, \dots, n. \quad \#$$

BERNOULLI TRIALS & RELATED RV (1 2)

BERNOULLI TRIALS (1.2.1)

1. A "Bernoulli trial" focuses on a particular random experiment with only two possible outcomes: success or failure.
2. We call the random variables and the experiment obtained from Bernoulli trials as "Bernoulli random variables" and a "Bernoulli experiment" respectively.

BERNOULLI RV (1.2.2)

1. In particular, if B is a Bernoulli rv:
   ① then $P[B = \text{Success}]$, or $P(B)$, is equal to $P(B) = p$ (where p = probability of success); and
   ② $P[B = \text{Failure}]$, or $P(B^c)$, is equal to $P(B^c) = 1-p$.
2. Thus, the pmf of B is $f: \{0,1\} \to [0,1]$ by $f(0) = 1-p$ & $f(1) = p$, or equivalently by $f(x) = p^x (1-p)^{1-x}$ $\forall x \in \{0,1\}$.

BERNOULLI SEQUENCE (1.2.3)

1. A "Bernoulli sequence" occurs when
   ① we repeat a Bernoulli trial many times;
   ② the results are all independent; and
   ③ the success probability p stays the same.

BINOMIAL DISTRIBUTION: $X \sim \text{Binomial}(n,p)$ / $X \sim \text{Bin}(n,p)$ (1.2.4)

1. Let X be the rv equal to the number of successes after repeating a Bernoulli trial n times independently, with probability of success p.
   Then, we say X follows a binomial distribution, and write $X \sim \text{Binomial}(n,p)$.
2. In this case, the pmf of X is equal to $f: \{0, \dots, n\} \to [0,1]$ by $f(k) = C_n^k p^k (1-p)^{n-k}$ $\forall 0 \leq k \leq n$.

GEOMETRIC DISTRIBUTION: $X \sim \text{Geometric}(p)$ / $X \sim \text{Geo}(p)$ (1.2.5)

1. Repeat independent Bernoulli trials, with success probability p, until a trial is successful.
   Let the rv X be equal to the number of failures before the success was reached.
   Then, we say X follows a geometric distribution, and write $X \sim \text{Geometric}(p)$.
2. In this case, the pmf of X is equal to $f: \mathbb{N} \to [0,1]$ by $f(k) = (1-p)^k p$ $\forall k \in \mathbb{N}$.
3. Note that
   ① $P(X \geq n) = (1-p)^n$ $\forall n \in \mathbb{N}$; and

   Proof. $P(X \geq n) = \sum_{k=n}^{\infty} (1-p)^k p = (1-p)^n p \sum_{k=0}^{\infty} (1-p)^k = (1-p)^n p \left(\frac{1}{p}\right) = (1-p)^n$. ∎

   ② $P(X \geq m+n \mid X \geq n) = P(X \geq m)$ $\forall m, n \in \mathbb{N}$ (the memory-less property).

   Proof. $P(X \geq m+n \mid X \geq n) = \frac{P(X \geq m+n \,\cap\, X \geq n)}{P(X \geq n)} = \frac{P(X \geq m+n)}{P(X \geq n)} = \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = P(X \geq m)$. ∎

NEGATIVE BINOMIAL DISTRIBUTION:

$X \sim$ Negative Binomial (k,p) / $X \sim NB(k,p)$ (1.2.6)

[1] Repeat independent Bernoulli trials, with success probability p, until the k^{th} success is reached.

Let the rv X be the number of failures before the k^{th} success.

Then, we say X follows a negative binomial distribution, and write $X \sim$ Negative Binomial (k,p).

[2] In this case, the pmf of X is equal to

$f: \mathbb{N} \to [0,1]$ by $f(n) = C^{n}_{n+k-1} p^k (1-p)^n \quad \forall n \in \mathbb{N}$.

Proof. See that

$$P[X=n] = P[\text{having } n \text{ failures before } k^{th} \text{ success}]$$
$$= P[n \text{ failures } \& \ k-1 \text{ successes, followed by } k^{th} \text{ success}]$$
$$= \frac{(n+k-1)!}{n!\,(k-1)!}(1-p)^n (p)^{k-1} \cdot p$$

$$\therefore P[X=n] = C^{n}_{n+k-1}(1-p)^n p^k. \quad \square$$

HYPERGEOMETRIC DISTRIBUTION:

$X \sim$ Hypergeometric (N, M, n) / $X \sim Hyp(N,M,n)$ (1.3)

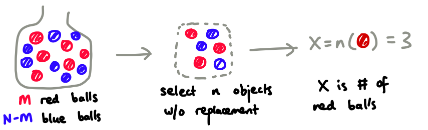
[1] Suppose we have a collection of N objects; M of one type, and $N-M$ of another (distinct) type.

Randomly select n objects without replacement, where $n \leq \min\{M, N-M\}$.

Let the rv X be the number of objects of the first type in these n objects.

Then, we say X follows a "hypergeometric distribution", and write

$$X \sim \text{Hypergeometric}(N, M, n).$$



[2] In this case, the pmf of X is equal to

$f: \{0, \dots, n\} \to [0,1]$ by $f(k) = \dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad \forall k=0,\dots,n.$

VANDERMONDE'S IDENTITY: $\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}$

[1] Let $n \leq M, N-M$.

Then necessarily

$$\binom{N}{n} = \sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k}.$$

POISSON DISTRIBUTION:

$X \sim$ Poisson (λ) / $X \sim Poi(\lambda)$ (1.4)

[1] In some observational studies, events happen over time or space.

We say such an event follows a Poisson process if the following conditions are satisfied:

① Events in non-overlap time intervals are independent; } independence

② $P[\geq 2 \text{ events in } [t, t+\Delta t)] = o(\Delta t)$, where $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$ and $\Delta t \ll t$; and } individuality

③ $P[\text{one event in } [t, t+\Delta t)] = \lambda \Delta t + o(\Delta t)$, $\lambda \in \mathbb{R}$. } homogeneity

Note that we call "λ" in ③ the "intensity parameter".

[2] Let the rv X be the number of events in $[0,t]$.

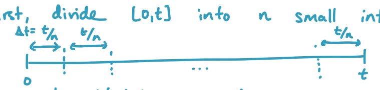
Then we say X follows a Poisson distribution, and write

$$X \sim \text{Poisson}(\lambda).$$

[3] In this case, the pmf of X is given by

$f: \mathbb{N} \to [0,1]$ by $f(x) = \dfrac{e^{-\lambda t}(\lambda t)^x}{x!}$.

Proof. First, divide $[0,t]$ into n small intervals:



Note that $\Delta t \to 0$ as $n \to \infty$.

Let the events

$B_1^{(n,x)}$ = there are x small intervals each with one event; &

$B_2^{(n)}$ = ≥ 1 small interval exists with two or more events.

Then, see that

$$P(B_1^{(n,x)}) = \binom{n}{x}\left(P(\text{one event in interval of length } \tfrac{t}{n} = \Delta t)\right)^x (1-p)^{n-x}$$

(by binomial distn)

$$= \binom{n}{x}\left(\lambda \tfrac{t}{n} + o(\tfrac{t}{n})\right)^x \left(1 - \lambda\tfrac{t}{n} - o(\tfrac{t}{n})\right)^{n-x}.$$

(by point ③ of defn)

Notice that since we want to consider infinitely small periods of time for our Poisson variable, we can deduce that

$$P(X=x) = \lim_{n\to\infty} P(B_1^{(n,x)})$$
$$= \lim_{n\to\infty}\left[\binom{n}{x}\left(\lambda\tfrac{t}{n} + o(\tfrac{t}{n})\right)^x\left(1-\lambda\tfrac{t}{n} - o(\tfrac{t}{n})\right)^{n-x}\right]$$
$$= \lim_{n\to\infty}\left[\frac{n!}{x!(n-x)!}\left(\lambda\tfrac{t}{n} + o(\tfrac{t}{n})\right)^x\left(1-\lambda\tfrac{t}{n} - o(\tfrac{t}{n})\right)^{n-x}\right]$$
$$= \lim_{n\to\infty}\left[\frac{1}{x!}\,\frac{n(n-1)\cdots(n-x+1)}{n^x}\left(\lambda t + n\,o(\tfrac{t}{n})\right)^x\left(1-\frac{\lambda t}{n} - o(\tfrac{t}{n})\right)^n \cdot \left(1-\frac{\lambda t}{n} - o(\tfrac{t}{n})\right)^{-x}\right]$$
$$= \frac{1}{x!}(1)(\lambda t)^x \lim_{n\to\infty}\left(1-\frac{\lambda t}{n}\right)^n (1)$$
$$= \frac{1}{x!}(\lambda t)^x e^{-\lambda t} \quad \left(\text{using the identity } e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n\right)$$

$$\therefore P(X=x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!},$$

as needed $\square$

Chapter 6:
Expectation and Variance

EXPECTED VALUE / EXPECTATION [OF A DISC RV] (1.1)

Let X be a drv, with $ran(X) = A$ & pmf $f(x)$.
Then, the "expectation" or "expected value" of X, denoted as "$E(X)$", is defined to be equal to

$$E(X) = \sum_{x \in A} x f(x). \quad (D1)$$

Note to calculate expectations, we need to:
① Identify the rv X involved;
② Find the pmf of X; and
③ Compute $E(X)$.

$X \sim \text{Bernoulli}(p) \Rightarrow E(X) = p$ (1.2.1)

Let $X \sim \text{Bernoulli}(p)$. Then necessarily $E(X) = p$.

Proof. $E(X) = \sum_{x \in \{0,1\}} x f(x)$
$= 0P(X^c) + 1P(X)$
$= p$. ∎

$X \sim \text{Binomial}(n,p) \Rightarrow E(X) = np$ (1.2.2)

Let $X \sim \text{Binomial}(n,p)$. Then necessarily $E(X) = np$.

Proof. $E(X) = \sum_{k=0}^{n} k P[X=k]$
$= \sum_{k=0}^{n} k\left(\binom{n}{k} p^k (1-p)^{n-k}\right)$
$= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$
$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$
$= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^k (1-p)^{n-k-1}$
$= np(1)$ (by Bin Formula)
$\therefore E(X) = np$. ∎

$X \sim \text{Geometric}(p) \Rightarrow E(X) = \frac{1-p}{p}$ (1.2.3)

Let $X \sim \text{Geometric}(p)$. Then necessarily $E(X) = \frac{1-p}{p}$.

Proof. $E(X) = \sum_{k=0}^{\infty} k \cdot P(X=k)$
$= \sum_{k=0}^{\infty} k(1-p)^k p$
$= p(1-p) \sum_{k=1}^{\infty} k(1-p)^{k-1}$.

Recall the identity
$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \dots + kx^{k-1} + \dots, \quad |x|<1$

Since $|1-p| < 1$, thus
$\frac{1}{p^2} = \frac{1}{(1-(1-p))^2} = 1 + 2(1-p) + 3(1-p)^2 + \dots = \sum_{k=1}^{\infty} k(1-p)^{k-1}$,

and so
$E(X) = p(1-p)\left(\frac{1}{p^2}\right) = \frac{1-p}{p}$. ∎

$X \sim NB(k,p) \Rightarrow E(X) = \frac{k(1-p)}{p}$ (1.2.4)

Let $X \sim NB(k,p)$. Then necessarily $E(X) = \frac{k(1-p)}{p}$.

Proof. $E(X) = \sum_{n=0}^{\infty} n \cdot P(X=n)$
$= \sum_{n=1}^{\infty} n \binom{n+k-1}{n} p^k (1-p)^n$
$= \sum_{n=1}^{\infty} n \frac{(n+k-1)!}{n!(k-1)!} (1-p)^n p^k$
$= \sum_{n=1}^{\infty} k \frac{(1-p)}{p} \cdot \frac{((n-1)+(k+1)-1)!}{(n-1)!\,k!} (1-p)^{n-1} p^{k+1}$
$= \frac{k(1-p)}{p} \sum_{n=1}^{\infty} \binom{(n-1)+(k+1)-1}{k+1} (1-p)^{n-1} p^{k+1}$
$= \frac{k(1-p)}{p} \sum_{n=0}^{\infty} \binom{n+(k+1)-1}{k+1} (1-p)^n p^{k+1}$
$= \frac{k(1-p)}{p}(1)$
$\therefore E(X) = \frac{k(1-p)}{p}$. ∎

$X \sim Hyp(N,M,n) \Rightarrow E(X) = \frac{nM}{N}$ (1.2.5)

Let $X \sim Hyp(N,M,n)$. Then necessarily $E(X) = \frac{nM}{N}$.

Proof. $E(X) = \sum_{k=0}^{n} k \cdot P(X=k)$
$= \sum_{k=1}^{n} k \cdot \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$
$= \frac{n!(N-n)!}{N!} \sum_{k=1}^{n} k \cdot \frac{M!}{k!(M-k)!} \frac{(N-M)!}{(n-k)!(N-M-n+k)!}$
$= \frac{n!(N-n)!}{N!} \sum_{k=1}^{n} M \cdot \frac{(M-1)!}{(k-1)!(M-k)!} \frac{(N-M)!}{(n-k)!(N-M-n+k)!}$
$= \frac{M(n!)(N-n)!}{N!} \sum_{k=1}^{n} \binom{M-1}{k-1}\binom{N-M}{n-k}$
$= M \frac{n!(N-n)!}{N!} \sum_{k=0}^{n-1} \binom{M-1}{k}\binom{N-M}{n-k-1}$
$= M \frac{n!(N-n)!}{N!} \binom{N-1}{n-1}$ (by Vandermonde's Identity)
$= M \frac{n!(N-n)!}{N!} \cdot \frac{(N-1)!}{(n-1)!(N-n)!}$

$\therefore E(X) = \frac{Mn}{N}$. ∎

$X \sim Po(\mu) \Rightarrow E(X) = \mu$ (1.2.6)

Let $X \sim Po(\mu)$. Then necessarily $E(X) = \mu$.

Proof. $E(X) = \sum_{n=0}^{\infty} n \cdot \frac{e^{-\mu}\mu^n}{n!}$
$= \sum_{n=1}^{\infty} \frac{e^{-\mu}\mu^n}{(n-1)!}$
$= \mu \sum_{n=0}^{\infty} \frac{e^{-\mu}\mu^n}{n!}$
$= \mu(1) = \mu$. ∎

$E[g(X)] = \sum_{x \in A} g(x) f(x)$ ⟪THE LAW OF THE UNCONSCIOUS STATISTICIAN⟫ (1.3)

Let g be a function on the drv X, which has range A and pmf X.
Then necessarily $E[g(X)] = \sum_{x \in A} g(x) f(x)$.

Proof. Let $Y = g(X)$, and let $D_y = \{x \in X : g(x) = y\}$, and let $B = ran(Y)$.
Then
$P[Y=y] = P[g(X)=y] = \sum_{x \in D_y} P[X=x]$.
Hence
$E(Y) = \sum_{y \in B} y \cdot P[g(X)=y]$
$= \sum_{y \in B} y \cdot \sum_{x \in D_y} P[X=x]$
$= \sum_{y \in B} \sum_{x \in D_y} g(x) P[X=x]$
$\therefore E(Y) = E(g(X)) = \sum_{x \in A} g(x) f(x)$. ∎

$E[ag(X)+b] = aE[g(X)] + b$; $E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$ ⟪LINEAR PROPERTIES OF EXPECTATION⟫

Let g, g_1, g_2 be functions on the drv X, and let $a, b \in \mathbb{R}^+$.
Then necessarily
① $E[ag(X)+b] = aE[g(X)] + b$; and
② $E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$.

Proof. Follows almost directly from LOUS. ∎

VARIANCE (1.4)

1. Let X be a drv. Then, the "variance" of X, denoted as "$\mathrm{Var}(X)$" or "σ^2", is defined to be equal to
$$\mathrm{Var}(X) = E[(X - E[X])^2].$$
* we use "μ" to denote "$E(X)$", so
$$\mathrm{Var}(X) = E[(X-\mu)^2].$$

2. $\mathrm{Var}(X)$ is used to measure the "variability" of a sample, ie how "concentrated" the data is.

3. The "standard deviation" of X, denoted as "σ", is defined to be
$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

$\mathrm{Var}(X) = \sum_{x \in A} (x - E(X))^2 f(x) = E(X^2) - [E(X)]^2$

Let X be a drv, with $\mathrm{ran}(X) = A$ & pmf f.
Then necessarily

① $\mathrm{Var}(X) = \sum_{x\in A} (x - E(X))^2 f(x)$; and

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

Proof. By def^n,
$$\mathrm{Var}(X) = E[(x-E(X))^2] = \sum_{x\in A}(x-E(X))^2 f(x) \quad \text{(by LOUS)},$$
showing ①.
Then
$$\begin{aligned}\sum_{x\in A}(x-E(X))^2 f(x) &= \sum_{x\in A}(x^2 - 2xE(X) + E(X)^2) f(x)\\ &= \sum_{x\in A} x^2 f(x) - 2E(X)\sum_{x\in A} x + E(X)^2\\ &= \sum_{x\in A} x^2 f(x) - 2E(X)E(X) + E(X)E(X)\\ &= E(X^2) - [E(X)]^2,\end{aligned}$$
showing ②. ∎

PROPERTIES OF VARIANCE

Let X be a drv. Note the following:

① $\mathrm{Var}(X) \geq 0$;

② $E(X^2) \geq (E(X))^2$;

③ $\mathrm{Var}(X) = 0 \iff P(X=c) = 1$ for some constant c; and

④ $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

Proof. By def^n of variance,
$$\mathrm{Var}(X) = \sum_{x\in A}(x - E(X))^2 f(x) = \sum_{x\in A}(x-E(X))^2 P[X=x],$$
and as $P[X=x],\ (x-E(X))^2 \geq 0 \quad \forall x \in A$, ① follows.
Thus
$$E(X^2) - (E(X))^2 \geq 0 \implies E(X)^2 \geq (E(X))^2,$$ showing ③.
Next,
$$\begin{aligned}\mathrm{Var}(X) = 0 &\iff \sum_{x\in A} \underbrace{(x-\mu)^2}_{\geq 0}\underbrace{f(x)}_{>0} = 0\\ &\iff (x-\mu)^2 = 0 \quad \forall x\in A\\ &\iff x = \mu \quad \forall x \in A \quad (\text{ie } A = \{\mu\})\\ &\iff P[X=\mu] = 1,\end{aligned}$$
showing ③;
and
$$\begin{aligned}\mathrm{Var}(aX+b) &= E[(aX+b)^2] - (E(aX+b))^2\\ &= a^2E(X^2) + 2abE(X) + b^2 - (aE(X)+b)^2\\ &= a^2E(X^2) + 2abE(X) + b^2 - a^2E(X)^2 - 2abE(X) - b^2\\ &= a^2E(X^2) - a^2E(X)^2\\ &= a^2\mathrm{Var}(X),\end{aligned}$$
showing ④. ∎

$X \sim \mathrm{Bernoulli}(p) \Rightarrow \mathrm{Var}(X) = p(1-p)$ (1.5.1)

Let $X \sim \mathrm{Bernoulli}(p)$. Then necessarily $\mathrm{Var}(X) = p(1-p)$.

Proof.
$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - (E(X))^2\\ &= \sum_{x=0}^{1} x^2 P[X=x] - p^2\\ &= p - p^2\end{aligned}$$
$\therefore \mathrm{Var}(X) = p(1-p)$. ∎

$X \sim \mathrm{Binomial}(n,p) \Rightarrow \mathrm{Var}(X) = np(1-p)$ (1.5.2)

Let $X \sim \mathrm{Binomial}(n,p)$. Then necessarily $\mathrm{Var}(X) = np(1-p)$.

Proof. First, see that
$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - (E(X))^2\\ &= E(X(X-1)) - (E(X))^2 + E(X).\end{aligned}$$
Then,
$$\begin{aligned}E(X(X-1)) &= \sum_{k=0}^{n} k(k-1)P[X=k]\\ &= \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\ &= \sum_{k=2}^{n} n(n-1)\cdot\frac{(n-2)!}{(k-2)!(n-k)!}p^k(1-p)^{n-k}\\ &= n(n-1)p^2\sum_{k=0}^{n-2}\frac{(n-2)!}{k!(n-k-2)!}p^k(1-p)^{n-2-k}\\ &= n(n-1)p^2(1)\end{aligned}$$
$\therefore E(X(X-1)) = n(n-1)p^2$.

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

$X \sim \mathrm{Geometric}(p) \Rightarrow \mathrm{Var}(X) = \frac{(1-p)^2}{p^2} + \frac{1-p}{p}$

Let $X \sim \mathrm{Geometric}(p)$. Then necessarily $\mathrm{Var}(X) = \frac{(1-p)^2}{p^2} + \frac{1-p}{p}$.

Proof. See that
$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - (E(X))^2\\ &= E(X(X-1)) - (E(X))^2 + E(X).\end{aligned}$$
Then
$$\begin{aligned}E(X(X-1)) &= \sum_{n=0}^{\infty} n(n-1)\cdot P[X=n]\\ &= \sum_{n=2}^{\infty} n(n-1)\cdot(1-p)^n p\\ &= p(1-p)^2[(1\times2) + (2\times3)(1-p) + (3\times4)(1-p)^2 + \cdots]\\ &= 2p(1-p)^2[1 + 3(1-p) + 6(1-p)^2 + \cdots]\\ &= 2p(1-p)^2\left(\frac{1}{(1-(1-p))^3}\right)\end{aligned}$$
$\therefore E(X(X-1)) = \frac{2(1-p)^2}{p^2}$.

Thus
$$\begin{aligned}\mathrm{Var}(X) &= E(X(X-1)) - (E(X))^2 + E(X)\\ &= \frac{2(1-p)^2}{p^2} - \left(\frac{1-p}{p}\right)^2 + \left(\frac{1-p}{p}\right)\\ &= \left(\frac{1-p}{p}\right)^2 + \left(\frac{1-p}{p}\right).\end{aligned}$$ ∎

$X \sim \mathrm{Poisson}(\mu) \Rightarrow \mathrm{Var}(X) = \mu$

Let $X \sim \mathrm{Poisson}(\mu)$. Then necessarily $\mathrm{Var}(X) = \mu$.

Proof. $\mathrm{Var}(X) = E(X(X-1)) - (E(X))^2 + E(X)$.
Then
$$\begin{aligned}E(X(X-1)) &= \sum_{n=0}^{\infty} n(n-1)P[X=n]\\ &= \sum_{n=2}^{\infty} n(n-1)\frac{e^{-\mu}\mu^n}{n!}\\ &= \mu^2\sum_{n=2}^{\infty}\frac{e^{-\mu}\mu^{n-2}}{(n-2)!}\\ &= \mu^2\sum_{n=0}^{\infty}\frac{e^{-\mu}\mu^n}{n!}\\ &= \mu^2(1) = \mu^2.\end{aligned}$$
Hence
$$\begin{aligned}\mathrm{Var}(X) &= E(X(X-1)) - (E(X))^2 + E(X)\\ &= \mu^2 - (\mu)^2 + \mu\\ &= \mu.\end{aligned}$$ ∎