Probability In Statistics

Shuai Wei

October 19, 2021

Contents

Chapter 1

Data Analysis

1.1 Three approaches to compute the sample mean $\overline{x}$

Assume we have a set of data points whose values are $x_1, \ldots, x_n$, respectively. Out of them, assume $v_1, \ldots, v_m$ are the m different values.

(a) Usual approach.

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}.$$

(b) Frequency approach. Assume $\omega_1, \ldots, \omega_m$ are the frequencies for $v_1, \ldots, v_m$, repectively. Then

$$\begin{aligned}\overline{x} &= \frac{v_1\omega_1 + \cdots + v_m\omega_m}{\omega_1 + \cdots + \omega_m} \\ &= \frac{\sum_{j=1}^m v_j\omega_j}{\sum_{j=1}^m \omega_j}.\end{aligned}$$

(c) Relative frequency approach. Assume $f_1, \ldots, f_m$ are the relative frequencies for $v_1, \ldots, v_m$, repectively. Then

$$\begin{aligned}\overline{x} &= v_1 f_1 + \cdots + v_m f_m \\ &= \sum_{j=1}^m v_j f_j.\end{aligned}$$

Chapter 2

Probability Rule

Let A and B be two events throughout.

2.1 Independence testing

We have

$$\begin{aligned} A \text{ and } B \text{ are indepdendent} &\iff P(A \text{ and } B) = P(A)P(B) \\ &\iff \frac{P(A \text{ and } B)}{P(A)} = P(B) \\ &\iff P(B|A) = P(B) \end{aligned}$$

$$\begin{aligned} A \text{ and } B \text{ are indepdendent} &\iff P(A \text{ and } B) = P(A)P(B) \\ &\iff \frac{P(A \text{ and } B)}{P(B)} = P(A) \\ &\iff P(A|B) = P(A) \end{aligned}$$

Hence

$$P(B|A) = P(B) \iff \underbrace{A \text{ and } B \text{ are independent}}_{P(A \text{ and } B)=P(A)P(B)} \iff P(A|B) = P(A).$$

2.2 Addition rule extensions

In general, we have the following 3 formulas to compute $P(A \text{ or } B)$.

$$\begin{aligned} P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) \\ &= P(A) + P(B) - P(A)P(B|A) \\ &= P(A) + P(B) - P(B)P(A|B). \end{aligned}$$

(a) In table form

$$\begin{aligned} P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) \\ &= \frac{\sharp\, A}{\sharp\,(\text{total observations})} + \frac{\sharp\, B}{\sharp\,(\text{total observations})} - \frac{\sharp\,(A \cap B)}{\sharp\,(\text{total observations})} \\ &= \frac{\sharp\, A + \sharp\, B - \sharp\,(A \cap B)}{\sharp\,(\text{total observations})}. \end{aligned}$$

(b) If A and B are mutually exclusive, then because $P(A \text{ and } B) = 0$,

$$P(A \text{ or } B) = P(A) + P(B).$$

(c) If A and B are independent, then because $P(A \text{ and } B) = P(A)P(B) = P(A)P(B|A) = P(B)P(A|B)$,

$$P(A \text{ or } B) = P(A) + P(B) - P(A)P(B).$$

2.3 Sequential events

Let A_1 and B_2 be sequential events such that A_1 happens first and then B_2 happens. Then always conditioning on A_1, we get

$$P(A_1B_2) = P(A_1 \text{ and } B_2) = P(A_1)P(B_2|A_1),$$

where we compute $P(B_2|A_1)$ directly, i.e., not use this formula $P(B_2|A_1) = \frac{P(A_1 \text{ and } B_2)}{P(A_1)}$.

2.4 Mutually exclusive events v.s. independent events

If you are interested in investigating more about the relationship between that two events are mutually exclusive and and that two events are independent, have a look at the following proposition.

Proposition 2.1. (a) Asume $P(A) \neq 0$ and $P(B) \neq 0$.

(1) If A and B are mutually exclusive, i.e., $P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B)$, then A and B are **not** independent. In other words, if A and B are independent, then A and B are **not** mutually exclusive.

(2) If A and B are **not** mutually exclusive, then whether A and B are independent or not depends on that $P(A \text{ and } B)$ equals to $P(A)P(B)$ or not.

(b) Assume $P(A) = 0$ or $P(B) = 0$, then since $P(A \cap B) \leq P(A)$ and $P(A \cap B) \leq P(B)$, we have

$$0 \leq P(A \text{ and } B) = P(A \cap B) \leq \min\{P(A), P(B)\} = 0 = P(A)P(B),$$

so $P(A \text{ and } B) = 0 = P(A)P(B)$, hence A and B are independent.

Chapter 3

Discrete Probability Distribution

The following example tells you more about the probability distribution of a discrete random variable in details.

Example 3.1. Let X be the number of heads observed in 3 tossing of a (fair) coin. So X takes values in the set $\{0, 1, 2, 3\}$. Let S be the sample space {HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}. Then the random variable X is a function from S to $\{0, 1, 2, 3\}$ by definition, which is defined by

$$\begin{aligned} X : S &\longrightarrow \{0, 1, 2, 3\} \\ \text{HHH} &\longmapsto 3 \\ \text{HHT} &\longmapsto 2 \\ \text{HTH} &\longmapsto 2 \\ \text{THH} &\longmapsto 2 \\ \text{HTT} &\longmapsto 1 \\ \text{THT} &\longmapsto 1 \\ \text{TTH} &\longmapsto 1 \\ \text{TTT} &\longmapsto 0 \end{aligned}$$

Let X^{-1} be the inverse map of X. Then for example,

$$X^{-1}(1) = \{s \in S : X(s) = 1\} = \{\text{HTT,THT,TTH}\}.$$

Note that $P(X = 1)$ is the probability of all the sample points s satisfying $X(s) = 1$, i.e.,

$$P(X = 1) = P(\{s \in S : X(s) = 1\}) = P(\{\text{HTT,THT,TTH}\}).$$

So actually, $P(X = 1) = P\big(X^{-1}(1)\big)$. Hence

$$
\begin{aligned}
\underbrace{P(X=1)}_{P(1)} &= P(\{\text{HTT,THT,TTH}\}) \\
&= P(\text{HTT}) + P(\text{THT}) + P(\text{TTH}) \\
&\quad \text{(b/c each tossing is independent to each other.)} \\
&= P(\text{H})P(\text{T})P(\text{T}) + P(\text{T})P(\text{H})P(\text{T}) + P(\text{T})P(\text{T})P(\text{H}) \\
&\quad \Big(\text{b/c coin is fair, } P(\text{H}) = \frac{1}{2} = P(\text{T}).\Big) \\
&= \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} \\
&= \Big(\frac{1}{2}\Big)^3 + \Big(\frac{1}{2}\Big)^3 + \Big(\frac{1}{2}\Big)^3 \\
&= \frac{3}{8}.
\end{aligned}
$$

Similarly, $\underbrace{P(X=0)}_{P(0)} = \frac{1}{8} = \underbrace{P(X=3)}_{P(3)}$ and $\underbrace{P(X=2)}_{P(2)} = \frac{3}{8}$.

Theorem 3.2. *Let X be a discrete random variable. If $a \neq b$, then the two events $\{X = a\}$ and $\{X = b\}$ are mutually exclusive.*

Proof. Suppose not. Then there exists $s \in \{X = a\} \cap \{X = b\}$. So $X(s) = a \neq b = X(s)$, implying that the function X is not well-defined, a contradiction. □

In Example 3.1, we have

$$
P(X \leq 1) = P(X = 0 \text{ or } X = 1) = \underbrace{P(X=0)}_{P(0)} + \underbrace{P(X=1)}_{P(1)} = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}.
$$

where the second equality follows from that $\{X = 0\}$ and $\{X = 1\}$ are mutually exclusive by theorem 3.2.

3.1 Binomial Distribution

Example 3.3. Let X be the number of heads observed in 3 tosses of an unfair coin with $p = P(H) = 0.3$. Then X is a binomial random variable with $n = 3$ and $p = 0.3$ and sample space

$$
\mathcal{S} = \{\text{HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}\}.
$$

Since 3 tosses (3 identical trials) in one experiment are independent, we have

$$\begin{aligned}
P(X=0) &= P(\mathrm{TTT}) = P(\mathrm{T})P(\mathrm{T})P(\mathrm{T}) = (1-0.3)^3,\\
P(X=1) &= P(\{\mathrm{HTT},\ \mathrm{THT},\ \mathrm{TTH}\}) = P(\mathrm{HTT}) + P(\mathrm{THT}) + P(\mathrm{TTH})\\
&= P(\mathrm{H})P(\mathrm{T})P(\mathrm{T}) + P(\mathrm{T})P(\mathrm{H})P(\mathrm{T}) + P(\mathrm{T})P(\mathrm{T})P(\mathrm{H})\\
&= 3P(\mathrm{H})P(\mathrm{T})^2 = 3\times 0.3\times(1-0.3)^2,\\
P(X=2) &= P(\{\mathrm{HHT},\ \mathrm{HTH},\ \mathrm{THH}\}) = P(\mathrm{HHT}) + P(\mathrm{HTH}) + P(\mathrm{THH})\\
&= P(\mathrm{H})P(\mathrm{H})P(\mathrm{T}) + P(\mathrm{H})P(\mathrm{T})P(\mathrm{H}) + P(\mathrm{T})P(\mathrm{H})P(\mathrm{H})\\
&= 3P(\mathrm{H})^2P(\mathrm{T}) = 3\times 0.3^2\times(1-0.3),\\
P(X=3) &= P(\mathrm{HHH}) = P(\mathrm{H})P(\mathrm{H})P(\mathrm{H}) = 0.3^3.
\end{aligned}$$

In this example, we can see that the **disjoint** union

$$\begin{aligned}
&\{X=0\}\sqcup\{X=1\}\sqcup\{X=2\}\sqcup\{X=3\}\\
&=\{\mathrm{TTT}\}\sqcup\{\mathrm{HTT,THT,TTH}\}\sqcup\{\mathrm{HHT,HTH,THH}\}\sqcup\{\mathrm{HHH}\}\\
&=\{\mathrm{TTT,HTT,THT,TTH,HHT,HTH,THH,HHH}\}\\
&=\{\mathrm{HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}\}\\
&=\mathcal{S}.
\end{aligned}$$

We say that these events $\{X=0\}$, $\{X=1\}$, $\{X=2\}$ and $\{X=3\}$ forms a **partition** of the sample space $\mathcal{S}$.

Theorem 3.4 (Binomial theorem)**.** *We expand*

$$(x+y)^n = \sum_{i=0}^{n} {}_n\mathrm{C}_i x^i y^{n-i}.$$

Example 3.5. (a) Put $n=3$ in the Binomial theorem, we get

$$\begin{aligned}
(x+y)^3 &= \sum_{i=0}^{n} {}_3\mathrm{C}_i x^i y^{3-i} = {}_3\mathrm{C}_0x^0y^3 + {}_3\mathrm{C}_1x^1y^2 + {}_3\mathrm{C}_2x^2y^1 + {}_3\mathrm{C}_3x^3y^0\\
&= y^3 + 3xy^2 + 3x^2y + x^3\\
&= x^3 + 3x^2y + 3xy^2 + y^3.
\end{aligned}$$

(b) Put $x=1$ and $y=1$ in the Binomial theorem, we get

$$(1+1)^n = \sum_{i=0}^{n} {}_n\mathrm{C}_i 1^i 1^{n-i} = \sum_{i=0}^{n} {}_n\mathrm{C}_i,$$

i.e.,

$$2^n = \sum_{i=0}^{n} {}_n\mathrm{C}_i.$$

In particular, when $n=3$, we have

$$2^3 = \sum_{i=0}^{3} {}_3\mathrm{C}_i = {}_3\mathrm{C}_0 + {}_3\mathrm{C}_1 + {}_3\mathrm{C}_2 + {}_3\mathrm{C}_3 = 1+3+3+1 = 8.$$

Theorem 3.6 (Binomial formula)**.** *Let X be a binomial random variable with n, p. Then*

$$P(X = i) = {}_n\mathrm{C}_i \cdot p^i(1-p)^{n-i},\ i = 0, 1, \ldots, n.$$

Proof. Since the event $\{X = i\}$ consisting of all sample points which have exactly i sucesses in n identical trials, by counting rules, for $i = 0, 1, \ldots, n$, we have the size

$$\sharp\,\{X = i\} = {}_n\mathrm{C}_i.$$

For each sample point $s \in \{X = i\}$, there are i success and $n - i$ failure, then by the independence of n identical trials, we have

$$P(s) = P(S)^i P(F)^{n-i} = p^i(1-p)^{n-i}.$$

Since every sample point in $\{X = i\}$ has the same probability $P(s)$, we have

$$P(X = i) = P(\{X = i\}) = \sharp\,\{X = i\} \cdot P(s) = {}_n\mathrm{C}_i \cdot p^i(1-p)^{n-i}. \qquad \square$$

Remark. It is straightforword to show that $\{X = 0\}, \{X = 1\}, \ldots, \{X = n\}$ forms a partition of the sample space $\mathcal{S}$ of a binomial random variable X with n, p. So the size of $\mathcal{S}$ is

$$\begin{aligned}
\sharp\,\mathcal{S} &= \sharp\,\{X = 0\} + \sharp\,\{X = 1\} + \cdots + \sharp\,\{X = n\} \\
&= \sum_{i=0}^{n} \sharp\,\{X = i\} \\
&= \sum_{i=0}^{n} {}_n\mathrm{C}_i \\
&= 2^n,
\end{aligned}$$

where the last equality follows from Example 3.5(b). In particular, in Example 3.3, we have

$$\sharp\,\mathcal{S} = \sharp\{\mathrm{HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}\} = 2^3 = 8.$$

Example 3.7. Let X be the number of heads observed in 99 tosses of a fair coin.

(a) Is $\{X = 1\}$ and $\{X = 3\}$ independent?

(b) $P(X = 2) = P(X = 97)$?

(c) $\sum_{i=0}^{49} P(X = i) = 0.5$?

(a) No. $P(X = 1 \text{ and } X = 3) = 0$ but $P(X = 1)P(X = 3) \neq 0$.

(b) Yes. Because

$$\begin{aligned} P(X = 2) &= {}_{99}\mathrm{C}_2 \cdot 0.5^2(1 - 0.5)^{97} \\ &= {}_{99}\mathrm{C}_2 \cdot 0.5^{99} \\ &= {}_{99}\mathrm{C}_{97} \cdot 0.5^{99} \\ &= {}_{99}\mathrm{C}_{97} \cdot 0.5^{97}(1 - 0.5)^2 \\ &= P(X = 97). \end{aligned}$$

(c) Yes, because

$$P(X = i) = {}_{99}\mathrm{C}_i \cdot (0.5)^{99} = {}_{99}\mathrm{C}_{99-i} \cdot (0.5)^{99} = P(X = 99 - i),\ i = 0, \ldots, 49.$$

Example 3.8. Which one of the followings is a binomial random variable?

(a) X is the number of 6 dots observed in 10 rolls of a dice.

(b) X is the number of days it rains in 2021.

(c) Little Corey is playing coin flipping game, let X be the number of tosses until he sees a tail.

(d) In Corey's playhouse, there are 10 empty boxes and 10 boxes each of which has 1 candy. Each time, he randomly opens a box that hasn't been opened by him, and eats the candy if there is one. Let X be the number candies he eats after he opened 6 boxes randomly one by one.

(a) Yes, X is a binomial random variable with $n = 10$ and $p = 1/6$.

(b) No, p is not fixed.

(c) No, n is not fixed.

(d) No, $p = 10/20 = 1/2$ when he open a box the first time, but when he open a box the second time, we have $p = 10/19$ if the first time he picked an empty box or $p = 9/19$ if the first time he picked a box with candy.

Chapter 4

Continuous distribution

Let X be a continuous random variable with **probability density function** (pdf) $f(x)$. For any constant a with $a \in (-\infty, \infty)$, we have

$$P(X < a) = \int_{-\infty}^{a} f(x)dx.$$

Then we naturally get a function on $(-\infty, \infty)$, called **cumulative distribution function** (cdf) $F(x)$,

$$F(x) := \underbrace{P(X \leq x) = P(X < x)}_{X \text{ is continuous}} = \int_{-\infty}^{x} f(t)dt, \ x \in (-\infty, \infty).$$

If $f(x)$ is continuous, then by the fundamental theorem of calculus, we know that F is differentiable with

$$F'(x) = f(x),$$

i.e., $F(x)$ is the antiderivate of $f(x)$.

Remark. If X is a continuous random variable, then the pdf $f(x)$ of X are not necessarily continuous, for example,

$$f(x) = \begin{cases} \frac{1}{14} & \text{if } 1 \leq x \leq 15 \\ 0 & \text{otherwise.} \end{cases}$$

4.1 Convert a normal random variable into a standard normal random variable

Fact 4.1. Let X be a **normal** random variable with mean μ and standard deviation σ. Then $\frac{X-\mu}{\sigma}$ is a **standard normal** random variable.

Proof. Define $Z := \frac{X-\mu}{\sigma}$. Then we need to show that the pdf of Z is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$. Note that

$$\begin{aligned} F(z) &= P(Z \le z) \\ &= P(Z < z) \\ &= P\Big(\frac{X-\mu}{\sigma} < z\Big) \\ &= P(X < \sigma z + \mu). \end{aligned}$$

Since the pdf of X is $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, we have

$$\begin{aligned} F(z) &= P(X < \sigma z + \mu) \\ &= \int_{-\infty}^{\sigma z+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \end{aligned}$$

Again, from calculus course, the derivative

$$\begin{aligned} F'(z) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{((\sigma z+\mu)-\mu)^2}{2\sigma^2}} \frac{d}{dz}(\sigma z + \mu) \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\sigma z)^2}{2\sigma^2}} \sigma \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \end{aligned}$$

From Section 0.1, we have the pdf

$$f(z) = F'(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Thus, $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable. □

Let X be a **normal** random variable with mean μ and standard deviation σ.

(a) One goal is to find $P(X < x)$, $P(X > y)$ or $P(x < X < y)$ given some constants x and y.

(1)

$$\begin{aligned} P(X < x) &= P(X - \mu < x - \mu) \\ &= P\Big(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\Big) \\ &= P\Big(Z < \frac{x-\mu}{\sigma}\Big). \end{aligned}$$

By Fact 0.1, we have $Z := \frac{X-\mu}{\sigma}$ is a standard normal random variable. Also, we have $\frac{x-\mu}{\sigma}$ is the z-score, so from norm table, we can look up for the probability $P(Z < \frac{x-\mu}{\sigma})$. That's why we can do that process as the textbook tells us.

(2)

$$\begin{aligned}P(X > y) &= 1 - P(X < y)\\&= 1 - P\Big(\frac{X-\mu}{\sigma} < \frac{y-\mu}{\sigma}\Big)\\&= 1 - P\Big(Z < \frac{y-\mu}{\sigma}\Big).\end{aligned}$$

(3)

$$\begin{aligned}P(x < X < y) &= P(X < y) - P(X < x)\\&= P\Big(\frac{X-\mu}{\sigma} < \frac{y-\mu}{\sigma}\Big) - P\Big(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\Big)\\&= P\Big(Z < \frac{y-\mu}{\sigma}\Big) - P\Big(Z < \frac{x-\mu}{\sigma}\Big).\end{aligned}$$

We have the similar analysis for (2) and (3) as in (1). So from now on, if we are given that X is a normal random variable, then

$$\begin{aligned}P(X < x) &= P\Big(Z < \frac{x-\mu}{\sigma}\Big)\\P(X > y) &= 1 - P\Big(Z < \frac{y-\mu}{\sigma}\Big)\\P(x < X < y) &= P\Big(Z < \frac{y-\mu}{\sigma}\Big) - P\Big(Z < \frac{x-\mu}{\sigma}\Big),\end{aligned}$$

where all the probabilities on the right can be directly looked up from norm table since $\frac{x-\mu}{\sigma}$ and $\frac{y-\mu}{\sigma}$ are z-scores.

(b) Another goal is to find x given the probability $P(X < x)$. Define $Z := \frac{X-\mu}{\sigma}$. Then Z is a standard normal random variable. Set $z = \frac{x-\mu}{\sigma}$. So

$$P(Z < z) = P\Big(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\Big) = P(X - \mu < x - \mu) = P(X < x).$$

Since $P(X < x)$ is given, we know the probability $P(Z < z)$. Let's look up norm table inversely, then we get the z-score z. Since $z = \frac{x-\mu}{\sigma}$, we get $x = \sigma z + \mu$.

Note 4.2. Let Z be a standard normal random variable, then for any constant $z > 0$, since the pdf for Z is symmetric about 0, we have

$$P(Z > z) = P(Z < -z).$$

$$P(Z > -z) = P(Z < z).$$

You can draw the shaded ared under the stadard normal pdf curve to see this.

Chapter 5

Sampling Distribution for Sample Mean $\overline{X}$

Recall in Chapter 4(lecture notes), we have sample mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. In this chapter, $\bar{x}$ is usually a **real number** while $\overline{X}$ is the **random variable** sample mean (because different sample (of size n) has different mean).

On the one hand, assume the salaries of Clemson professors has a **normal** distribution with mean μ and standard deviation σ.

(a) The salary X of **a** randomly selected professor has a **normal** distribution with mean $\mu_X = \mu$ and standard deviation $\sigma_X = \sigma$.

(1) If x is known, then $P(X < x) = P(Z < \frac{x-\mu}{\sigma})$, which can be found from norm table based on the z-score $\frac{x-\mu}{\sigma}$; similarly we can compute $P(X > y)$ given y and $P(x < X < y)$ given x, y.

(2) If $P(X < x)$ is known, then we get the z-score z from norm table conversely based on the probability $P(X < x)$, so $x = \sigma z + \mu$ from the expression $z = \frac{x-\mu}{\sigma}$.

(b) The **average** salary $\overline{X}$ of n randomly selected professor(s) has a **normal** distribution with mean $\mu_{\overline{X}} = \mu$ and <u>standard error</u> $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$.

Remark. In part (b), if $n = 1$, then it reduces to part (a): The (average) salary $\overline{X}$ (X) of **1** randomly selected professor(s) has a **normal** distribution with mean $\mu_{\overline{X}} = \mu$ and standard error $\mu_{\overline{X}} = \frac{\sigma}{\sqrt{1}} = \sigma$.

On the other hand, assume the salaries of Clemson professors has an **unknown** distribution but known mean μ and standard deviation σ. Then the **average** salary $\overline{X}$ of n randomly selected professor(s) has mean $\mu_{\overline{X}} = \mu$ and <u>standard error</u> $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$. Moreover, if the sample size **satisfies** $n \geq 30$, then by Central Limit Theorem, the **average** salary $\overline{X}$ of n randomly selected professor has a **normal** distribution.

Assume $\overline{X}$ is normally distributed with $\mu_{\overline{X}} = \mu$ and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$.

(a) If $\bar{x}$ is known, then define $Z := \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, we have

$$P(\overline{X} < \bar{x}) = P(\overline{X} - \mu < \bar{x} - \mu) = P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right) = P\left(Z < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right),$$

which can be found from norm table based on the z-score $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$; similarly we can compute $P(\overline{X} > \bar{y})$ given $\bar{y}$ and $P(\bar{x} < \overline{X} < \bar{y})$ given $\bar{x}, \bar{y}$.

(b) If $P(\overline{X} < \bar{x})$ is known, then we get the z-score z from norm table based on the probability $P(\overline{X} < \bar{x})$, so $\bar{x} = \frac{\sigma}{\sqrt{n}} z + \mu$ from the expression $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

Chapter 6

Sampling Distribution for Sample Proportion $\hat{P}$

In this chapter, $\hat{p}$ is usually a **proportion** calculated from a **sample**, called **sample proportion**. When **I** regard sample proportion as a random variable or talk about the distribution of sample proportion, **I** use $\hat{P}$ to denote the **random variable**: sample proportion.

Teddy claims that he has an unfair coin in the sense that the probability of a head occuring is 0.64 in a flipping, in other words (in frequentists' view), if the coin is flipped many many times, there is approximate 64% of times that a head will be observed. Mary didn't believe this. So she flipped the coin for 100 times, and found that a head occured 52 times.

Assume Teddy didn't tell a lie. We know then the (sample) proportion $\hat{P}$ of a head occuring in 100 flipping of that coin is a random variable with mean $\mu_{\hat{P}} = p = 0.64$ and standard error $\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{100}} = \sqrt{\frac{0.64*0.36}{100}} = \frac{0.8*0.6}{10} = 0.048$. Since $np = 100 * 0.64 = 64 \geq 5$ and $n(1-p) = 100*(1-0.64) = 36 \geq 5$, we have $\hat{P}$ is a **norm** random variable (with mean $\mu_{\hat{P}} = 0.64$ and standard error $\sigma_{\hat{P}} = 0.048$).

Remark. That the sample proportion $\hat{P}$ is a random variable is because if Mary flipps the coin 100 times the second day, then a head may occur 51 times.

Assume $\hat{P}$ is normally distributed with $\mu_{\hat{P}} = p$ and $\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}}$.

(a) If $\hat{p}$ is known, then defining $Z := \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}}$, we have

$$P(\hat{P} < \hat{p}) = P(\hat{P} - p < \hat{p} - p) = P\left(\frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\right) = P\left(Z < \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\right),$$

which can be found from norm table based on the z-score $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$; similarly we can compute $P(\hat{P} > \hat{q})$ given $\hat{q}$ and $P(\hat{p} < \hat{P} < \hat{q})$ given $\hat{p}, \hat{q}$.

(b) If $P(\hat{P} < \hat{p})$ is known, then we get the z-score z from norm table conversely based on the probability $P(\hat{P} < \hat{p})$, so $\hat{p} = \sqrt{\frac{p(1-p)}{n}}z + p$ from the expression $z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$. (The lecture notes doesn't contain an example about this).

Example 6.1. Let's verify Teddy's claim from Mary's flipping. We have

$$P(\hat{P} < 0.52) = P\left(Z < \frac{0.52 - 0.64}{\sqrt{\frac{0.64(1-0.64)}{100}}}\right) = P\left(Z < \frac{-0.12}{0.048}\right) = 0.006.$$

So if Teddy's claim is correct, then the probability of observing 52 heads in 100 flipping is 0.006. It is too small to happen in practice, however, Mary had that result. The possible explanation is that Teddy's claim is **incorrect**.

Question 6.2. How about if Mary observed exact 64 heads in 100 flipping? Note that

$$P(\hat{P} < 0.64) = P\left(Z < \frac{0.64 - 0.64}{\sqrt{\frac{0.64(1-0.64)}{100}}}\right) = P(Z < 0) = 0.5.$$

I don't know if 0.5 can help Mary make conclusion or not. In Section 10.7 of lecture notes, you'll study how to reject someone's claim using hypothesis testing for sample proportion where exactly the same method will be used as in Example 6.1. Still you cannot support someone's claim.

Chapter 7

Confidence Interval for Proportions

Fact 7.1. If Z is a standard normal random variable, then $Z' := -Z$ is also a standard normal random variable.

Proof.

$$P(Z' < z) = P(-Z < z) = P(Z > -z) = 1 - P(Z < -z) = 1 - \int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

So the pdf of Z' is

$$f_{Z'}(z) = \frac{dP(Z' < z)}{dz} = 0 - \frac{1}{\sqrt{2\pi}} e^{-\frac{(-z)^2}{2}} \frac{d(-z)}{dz} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Thus, Z' is a standard normal random variable. $\square$

Why it is called 95%? From Chapter 8, if the population proportion p is known, we have that $Z := \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ is a standard normal random variable. Then by Fact 7.1, $Z' = \frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}}$ is also a standard normal random variable. In Chapter 9, the population proportion p is usually unkown, approximately (not accurately), we substitute $\hat{p}$ to estimate standard error $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Hence $Z' = \frac{p-\hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ is a standard normal random variable.

$$
\begin{aligned}
&P\left(\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)\\
&= P\left(-z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p - \hat{p} < z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)\\
&= P\left(-z_{0.025} < \frac{p-\hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < z_{0.025}\right)\\
&= P(-z_{0.025} < Z' < z_{0.025})\\
&= 1 - P(Z' < -z_{0.025}) - P(Z' > z_{0.025})\\
&= 1 - 2P(Z' > z_{0.025})\\
&= 1 - 2 \times 0.025\\
&= 1 - 0.05\\
&= 0.95.
\end{aligned}
$$

Warning 7.2. The most confusion part is however, we shouldn't say that "the probability that the population proportion p is between $\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ and $\hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is 0.95." Why? Probability statements are about random variables. The population proportion is a constant, not a random variable. It makes no sense to make a probability statement about a constant that does not change.

What is the **correct** understanding of confidence intervals?

Example 7.3. Assume true proportion of male is 0.5 in Clemson University.

(1) The first time we randomly pick 200 people from Clemson University, getting 97 males, so $\hat{p}_1 = \frac{97}{200} = 0.485$. The 95% confidence interval is $0.485 \pm z_{0.025}\sqrt{\frac{0.485*(1-0.485)}{200}} = [0.42, 0.55]$, which contains the true proportion $p = 0.5$. We are 95% confident that the population proportion of male in Clemson Univerity lies between 0.42 and 0.55. (The 95% confidence interval actually contains the true proportion $p = 0.5$).

(2) The second time we randomly pick 200 people from Clemson University, getting 110 males, so $\hat{p}_2 = \frac{110}{200} = 0.55$. The 95% confidence interval is $0.55 \pm z_{0.025}\sqrt{\frac{0.55*(1-0.55)}{200}} = [0.48, 0.62]$. We are 95% confident that the population proportion of male in Clemson Univerity lies between 0.48 and 0.62. (The 95% confidence interval actually contains the true proportion $p = 0.5$).

$\vdots$

(3) The 999th time we randomly pick 200 people from Clemson University, getting 120 males, so $\hat{p}_{999} = \frac{110}{200} = 0.6$. The 95% confidence interval is $0.6 \pm z_{0.025}\sqrt{\frac{0.6*(1-0.6)}{200}} = [0.53, 0.67]$. We are 95% confident that the population proportion of male in Clemson Univerity lies between 0.48 and 0.62. (The 95% confidence interval actually **doesn't** contain the true proportion $p = 0.5$).

$\vdots$

Remark. In the long run, 95% of all samples of size 200 from Clemson University will give a confidence interval that contains the true proportion $p = 0.5$. For example, if we repeatdly pick 200 people from Clemson University for 100000 times, then approximate 100000*95% = 95000 intervals will contains the population proportion $p = 0.5$. (In JMP Lab 5, you'll do some simulation work to see this.)