ST5201: Basic Statistical Theory
Tutorial

CHOI Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

14th November, 2017

Announcement

- Assignment 5 due today
- Final scheduled on 4th of December at 1:00 PM. The venue is LT31.
  - All contents we have learned from lecture 1 to lecture 9 are subject to final (mostly Chapter 1 to Chapter 9). For your information, lecture 10 and lecture 11 were review sessions.
  - You can bring one A4-size help sheet.
  - You can bring a non-programmable calculator, and you might actually need it.
  - Distribution tables (e.g., $Z-$table, χ^2-table, $t-$table) will be provided.

Simpson's Paradox

A black urn contains 5 red and 6 green balls, and a white urn contains 3 red and 4 green balls. You are allowed to choose an urn and then choose a ball at random from the urn. If you choose a red ball, you get a prize. Which urn should you choose to draw from?

Now consider another game in which a second black urn has 6 red and 3 green balls, and a second white urn has 9 red and 5 green balls. The rule is the same. Then which urn should you choose to draw from?

In the final game, the contents of the second black urn are added to the first black urn, and the contents of the second white urn are added to the first white urn. Again, you can choose which urn to draw from. Which should you choose?

Solution:

- For the first game, the probability of choosing a red ball from a black urn is $\frac{5}{5+6} \approx 0.455$ while that from a white urn is $\frac{3}{3+4} \approx 0.429$. Thus we would want to choose from a black urn
- For the second game, the probability of choosing a red ball from a black urn is $\frac{6}{6+3} \approx 0.667$ while that from a white urn is $\frac{9}{9+5} \approx 0.643$. Thus we would want to choose from a black urn
- For the final game, the probability of choosing a red ball from a black urn is $\frac{5+6}{11+9} = 0.550$ while that from a white urn is $\frac{3+9}{4+14} \approx 0.571$. Thus we would want to choose from a **white** urn

Solution continued

This counter-intuitive example is called *Simpson's Paradox*, that a trend appears in different groups of data but disappears or reverses when these groups are combined. If you check the pooled statistics only, you might be mislead to make wrong decisions/conclusions.

This phenomenon is often encountered in social science and medical science statistics. A well-known example is a study of gender bias among graduate school admissions to University of California, Berkeley. In 1973, the University of California-Berkeley was sued for sex discrimination. The numbers looked pretty incriminating: the graduate schools had just accepted 44% of male applicants but only 35% of female applicants. When researchers looked at the evidence, though, they uncovered this paradox. When they checked the ratio for each department, the admission rate of female applicants might be higher. The reason is that men more often applied to science departments, while women inclined towards humanities. Science departments require special technical skills but accept a large percentage of qualified applicants. In contrast, humanities departments only require a standard undergrad curriculum but have fewer slots.

Solution continued

Compare this example with our problem. **Female** applicants are balls in the **black urn**, and those who are accepted can be seen as the red balls. Male applicants are balls in the white urn, and those who are accepted can be seen as the red balls, again.

For the humanity department (the first case), there are more **female** applicants (more balls in the **black urn**), and they slightly tend to accept the **female** applicants (**black urn** has higher probability to select a red ball). However, they have fewer slots (the total number of red balls is small).

For the science department (the second case), there are less **female** applicants (less balls in the **black urn**), and they also tend to accept the **female** applicants (**black urn** has higher probability to select a red ball), and they have more slots (the total number of red balls is large). However, because there are many more male applicants, they will accept many more male applicants even thought the acceptance rate is lower (white urn have 9 red balls, which is more than 6 red balls in the **black urn**).

In total, it may come out the statistic male applicants are easier to be accepted (white urn has larger probability to select a red ball).

Statistics can fool us!

Capture/Recapture Method

The so-called capture/recapture method is sometimes used to estimate the size of a wildlife population. In detail, it is to capture several animals first, tag them, and then release them. After a while (allow sufficient time for the captured animals to distribute them into the population), capture the animals again, and count the number of tagged animals in these captured animals.

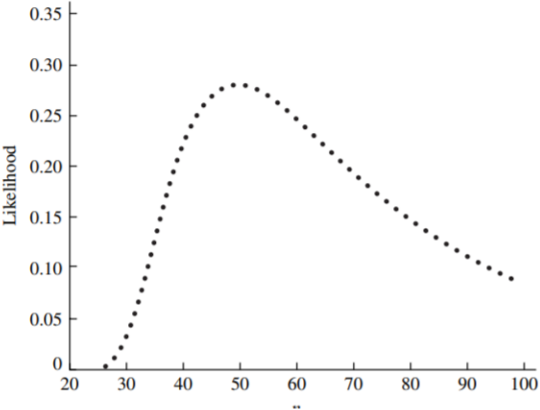
Suppose that 10 animals are captured, tagged, and released. On a later occasion, 20 animals are captured, and it is found that 4 of them are tagged. How large is the population?

Solution:

We assume that there are n animals in the population, of which 10 are tagged. If the 20 animals captured later are taken in such a way that all $\binom{n}{20}$ possible groups are equally likely (this is a big assumption), then the probability that 4 of them are tagged is (using the technique of the previous example)

$$\frac{\binom{10}{4}\binom{n-10}{16}}{\binom{n}{20}}$$

Clearly, n cannot be precisely determined from the information at hand, but it can be estimated. One method of estimation, called **maximum likelihood,** is to choose that value of n that makes the observed outcome most probable. (The method of maximum likelihood is one of the main subjects of a later chapter in this text.) The probability of the observed outcome as a function of n is called the **likelihood.** Figure 1.3 shows the likelihood as a function of n; the likelihood is maximized at $n = 50$.



Solution continued

Solution continued

To find the maximum likelihood estimate, suppose that, in general, t animals are tagged. Then, of a second sample of size m, r tagged animals are recaptured. We estimate n by the maximizer of the likelihood

$$L_n = \frac{\binom{t}{r}\binom{n-t}{m-r}}{\binom{n}{m}}$$

To find the value of n that maximizes L_n, consider the ratio of successive terms, which after some algebra is found to be

$$\frac{L_n}{L_{n-1}} = \frac{(n-t)(n-m)}{n(n-t-m+r)}$$

This ratio is greater than 1, i.e., L_n is increasing, if

$$\begin{aligned}
(n-t)(n-m) &> n(n-t-m+r) \\
n^2 - nm - nt + mt &> n^2 - nt - nm - nr \\
mt &> nr \\
\frac{mt}{r} &> n
\end{aligned}$$

Solution continued

Thus, L_n increases for $n < mt/r$ and decreases for $n > mt/r$; so the value of n that maximizes L_n is the greatest integer not exceeding mt/r.

Applying this result to the data given previously, we see that the maximum likelihood estimate of n is $\frac{mt}{r} = \frac{20 \cdot 10}{4} = 50$. This estimate has some intuitive appeal, as it equates the proportion of tagged animals in the second sample to the proportion in the population:

$$\frac{4}{20} = \frac{10}{n} \qquad \blacksquare$$

So, we have $n = \frac{10 \cdot 20}{4} = 50$

Bayes' Rule

Recently, there is a disease spreading. To make sure that Joe is healthy, Joe's doctor draws some of Joe's blood, and performs the test on his drawn blood. Here's the information that Joe's doctor knows about the disease and the diagnostic blood test:

- 1% of the people have the disease. That is, if D is the event that a randomly selected individual has the disease, then $P(D) = 0.01$.
- The sensitivity of the test is 0.95. That is, if a person has the disease, then the probability that the diagnostic blood test comes back positive is 0.95. That is, $P(T + |D) = 0.95$.
- The specificity of the test is 0.95. That is, if a person is healthy, then the probability that the diagnostic test comes back negative is 0.95. That is, $P(T - |H) = 0.95$.

Now the results indicate that the disease is present in Joe. What is the probability that Joe actually has the disease?

Solution:
In this problem, there are 4 events mentioned. Define the following events:

$$D = \{\text{Joe has the disease}\}, \quad H = \{\text{Joe is healthy}\},$$

$$T+ = \{\text{The blood test comes back positive}\}, \quad T- = \{\text{The blood test comes back negative}$$

With the conditions, we have that

$$P(D) = 0.01, \quad P(H) = 0.99, \quad P(T+|D) = 0.95, \quad P(T-|H) = 0.95.$$

The probability of interest is $P(H|T+)$, that given positive feedback from blood test, what is the probability of D. According to Bayes' Rule, we have that

$$P(D|T+) = \frac{P(T+|D)P(D)}{P(T+|H)P(H) + P(T+|D)P(D)} = \frac{.95 * .01}{(1-.95) * .99 + .95 * .01} \approx .161.$$

Given that the feedback is positive, the probability that Joe actually has disease is .161.

Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a prize; behind the others, nothing. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has nothing. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

For this question, we assume that the procedure is always fixed. It means that:

1. The host must open a door that was not picked by the contestant;
2. The host must open a door to reveal nothing behind it.

Monty Hall Problem–Solution

Solution:
Consider the events C_1, C_2 and C_3 indicating the prize is behind respectively door 1, 2 or 3. All these 3 events have probability 1/3.
The player initially choosing Door 1 is described by the event X_1. As the first choice of the player is independent of the position of the prize, also the conditional probabilities are $P(C_i|X_1) = 1/3$, $i = 1, 2, 3$. According to multiplication law,

$$P(C_i \cap X_1) = P(C_i \cap X_1)P(X_1) = 1/3 \times 1/3 = 1/9, \qquad i = 1, 2, 3.$$

The host opening door 3 is described by H_3. As it is required that the host must open a door not selected by the player and with nothing behind, so the conditional probabilities are

$$\begin{aligned} P(H_3|C_1 \cap X_1) &= 1/2; \\ P(H_3|C_2 \cap X_1) &= 1; \\ P(H_3|C_3 \cap X_1) &= 0. \end{aligned}$$

Monty Hall Problem–Solution

Solution continued
Then, if the player initially selects door 1, and the host opens door 3, the conditional probability that the prize is under door 1 is

$$
\begin{aligned}
P(C_1|H_3 \cap X_1) &= \frac{P(C_1 \cap H_3 \cap X_1)}{P(H_3 \cap X_1)} \\
&= \frac{P(H_3|C_1 \cap X_1)P(C_1 \cap X_1)}{P(H_3|C_1 \cap X_1)P(C_1 \cap X_1) + P(H_3|C_2 \cap X_1)P(C_2 \cap X_1) + P(H_3|C_3 \cap X_1)P(C_3 \cap} \\
&= \frac{1/2 \times 1/9}{1/2 \times 1/9 + 1 \times 1/9 + 0 \times 1/9} \\
&= 1/3.
\end{aligned}
$$

For the second equality, as $\{C_1, C_2, C_3\}$ is a partition of the sample space, we decompose the event $H_3 \cap X_1$ into $C_i \cap (H_3 \cap X_1)$, and calculate $P(C_i \cap (H_3 \cap X_1))$ by $P(H_3|C_i \cap X_1)P(C_i \cap X_1)$ with multiplication law.

Monty Hall Problem–Solution

Solution continued

Similarly, the conditional probability that the prize is under door 2 is

$$
\begin{aligned}
P(C_2|H_3 \cap X_1) &= \frac{P(C_2 \cap H_3 \cap X_1)}{P(H_3 \cap X_1)} \\
&= \frac{P(H_3|C_2 \cap X_1)P(C_2 \cap X_1)}{P(H_3|C_1 \cap X_1)P(C_1 \cap X_1) + P(H_3|C_2 \cap X_1)P(C_2 \cap X_1) + P(H_3|C} \\
&= \frac{1 \times 1/9}{1/2 \times 1/9 + 1 \times 1/9 + 0 \times 1/9} \\
&= 2/3.
\end{aligned}
$$

As $P(C_2|H_3 \cap X_1) > P(C_1|H_3 \cap X_1)$, the player should change the guess to Door 2 for a larger probability to win.

Gene Types

Assume that genes in an organism occur in pairs and that each member of the pair can be either of the types a or A. The possible genotypes of an organism are then AA, Aa, and aa (Aa and aA are equivalent). When two organisms mate, each independently contributes one of its two genes; either one of the pair is transmitted with probability .5.

a. Suppose that the genotypes of the parents are AA and Aa. Find the possible genotypes of their offspring and the corresponding probabilities.

b. Suppose that the probabilities of the genotypes AA, Aa, and aa are p, $2q$, and r, respectively, in the first generation. Find the probabilities in the second and third generations, and show that these are the same. This result is called the Hardy-Weinberg Law.

Solution:
(a). Let $A1 = \{$ Parent AA contributes a gene $A\}$ and $a1 = \{$ Parent AA contributes a gene $a\}$. Let $A2 =\{$ Parent Aa contributes a gene $A\}$ and $a2 = \{$ Parent Aa contributes a gene $a\}$.
As each parent independently contributes one of its two genes, we have $P(A1) = 1$, $P(a1) = 0$, $P(A2) = .5$, $P(a2) = .5$. So, the possible gene types for children are

$$\begin{aligned} P(\{AA\}) &= P(A1) \times P(A2) = 1 * .5 = .5; \\ P(\{Aa\}) &= P(A1) \times P(a2) + P(a1) \times P(A2) = 1 * .5 + 0 * .5 = .5; \\ P(\{aa\}) &= P(a1) \times P(a2) = 0 * .5 = 0. \end{aligned}$$

As a conclusion, their offspring has probability 1/2 to have genetype AA, and probability 1/2 to have genetype Aa.

Solution continued.
(b). In the first generation, the probabilities of the genotypes AA, Aa, and aa are p, $2q$, and r. So, the probability that an offspring in the second generation will receive gene A is

$$P(A) = p \times 1 + 2q \times .5 + r \times 0 = p + q,$$

and similarly, the probability that an offspring in the second generation will receive gene a is

$$P(a) = p \times 0 + 2q \times .5 + r \times 1 = q + r.$$

As each gene is passed independently, so for the offspring we have that

$$\begin{aligned} P(AA) &= P(A) \times P(A) = (p+q)^2, \\ P(Aa) &= \binom{2}{1} P(A) \times P(a) = 2(p+q)(q+r), \\ P(aa) &= P(a) \times P(a) = (q+r)^2. \end{aligned}$$

Solution continued.
Note that $p + 2q + r = 1$ according to the law of total probability, so we have

$$\begin{aligned} P(AA) &= [(p+q)^2 + (p+q)(q+r)]^2, && (1)\\ &= (p+q)^2[p+q+q+r]^2 && (2)\\ &= (p+q)^2. && (3) \end{aligned}$$

Similarly, there is

$$\begin{aligned} P(Aa) &= 2[(p+q)^2 + (p+q)(q+r)] \times [(p+q)(q+r) + (q+r)^2] = 2(p+q)(q+r),\\ P(aa) &= [(p+q)(q+r) + (q+r)^2]^2 = (q+r)^2. \end{aligned}$$

The probability is the same with the second generation.

Memoryless Property

If X is a geometric random variable, show that

$$P(X > n + k - 1 | X > n - 1) = P(X > k)$$

Memoryless Property–Solution

Solution. According to the definition of conditional probability, there is

$$P(X > n+k-1 | X > n-1) = \frac{P(\{X > n+k-1\} \cap \{X > n-1\})}{P(X > n-1)} = \frac{P(X > n+k-1)}{P(X > n-1)}.$$

According to pmf of geometric r.v., we have

$$P(X > m) = \sum_{x>m} P(X = x) = \sum_{x>m} q^{x-1}p = p \sum_{x \geq m+1} q^{x-1}.$$

According to geometric summation in mathematics, we have

$$\sum_{x \geq m+1} q^{x-1} = \frac{q^{m+1-1} \times 1}{1-q} = q^m/(1-q).$$

Introduce it into $P(X > m)$, and recall that $p = 1 - q$, we have that

$$P(X > m) = p \sum_{x \geq m+1} q^{x-1} = p \times \frac{q^m}{1-q} = q^m.$$

Therefore, we have $P(X > n+k-1) = q^{n+k-1}$, and $P(X > n-1) = q^{n-1}$. Hence,

$$P(X > n+k-1 | X > n-1) = \frac{P(X > n+k-1)}{P(X > n-1)} = \frac{q^{n+k-1}}{q^{n-1}} = q^k = P(X > k).$$

So the conclusion is proved. $\square$

Textbook Prob 3.2 and Extension

An urn contains 10 black balls, 8 white balls, and 7 red balls; and 6 balls are chosen without replacement.

a. Find the joint distribution of the numbers of black, white, and red balls in the sample
b. Find the joint distribution of the numbers of black and white balls in the sample
c. Find the marginal distribution of the number of white balls in the sample
d. Find the conditional pmf of the number of black balls in the sample given the number of white balls in the sample

Solution:

a. Let $X =\#$ of black balls in the sample, $Y = \#$ of white balls in the sample, and $Z = \#$ of red balls in the sample. According to the experiment, we know that X, Y and Z are non-negative integers, $X + Y + Z = 6$, and $X \leq \min\{6, 10\}$, $Y \leq \min\{6, 8\}$ and $Z \leq \min\{6, 7\}$.

According to the counting methods, the pmf for (X, Y, Z) is

$$p_{X,Y,Z}(x, y, z) = \frac{\binom{10}{x}\binom{8}{y}\binom{7}{z}}{\binom{25}{6}}, \qquad 0 \leq X, Y, Z \leq 6, X + Y + Z = 6,$$

and 0 otherwise.

b. For any pair (x, y), if there is $(X, Y) = (x, y)$, then it follows that $Z = 6 - x - y$. So, the joint distribution is

$$\begin{aligned} p_{X,Y}(x, y) &= \sum_{\text{all possible } z \text{ with } (x,y)} p_{X,Y,Z}(x, y, z) \\ &= p_{X,Y,Z}(x, y, 6 - x - y) = \frac{\binom{10}{x}\binom{8}{y}\binom{7}{6-x-y}}{\binom{25}{6}}, \quad 0 \leq X, Y \leq 6, X + Y \leq 6, \end{aligned}$$

and 0 otherwise.

Solution continued

c.

There are two ways to find out the marginal distribution of the number of white balls in the sample.

Approach 1. Use counting method to find out the marginal pmf directly. The event $\{Y = y\}$ is that there are y white balls in the sample and $6 - y$ non-white balls. There are $\binom{8}{y}\binom{10+7}{6-y}$ possible ways for this situation, and $\binom{25}{6}$ possible ways in total. So the pmf is

$$p_Y(y) = \frac{\binom{8}{y}\binom{17}{6-y}}{\binom{25}{6}}, \quad 0 \le y \le 6,$$

and 0 otherwise.

Solution continued

c. Approach 2. Summing up the joint density to find out the marginal density. For $Y = y$, possible x are $0, 1, \cdots, 6-y$ as $x + y \leq 6$. So the summation is

$$\begin{aligned}
p_Y(y) &= \sum_{x=0}^{6-y} p_{X,Y}(x,y) = \sum_{x=0}^{6-y} \frac{\binom{10}{x}\binom{8}{y}\binom{7}{6-x-y}}{\binom{25}{6}} \\
&= \frac{\binom{8}{y}}{\binom{25}{6}} \sum_{x=0}^{6-y} \binom{10}{x}\binom{7}{6-x-y} \\
&= \frac{\binom{8}{y}}{\binom{25}{6}} \sum_{x=0}^{6-y} \binom{10}{x}\binom{7}{6-x-y} \cdot \frac{\binom{17}{6-y}}{\binom{17}{6-y}} \\
&= \frac{\binom{8}{y}\binom{17}{6-y}}{\binom{25}{6}} \sum_{x=0}^{6-y} \frac{\binom{10}{x}\binom{7}{6-x-y}}{\binom{17}{6-y}}
\end{aligned}$$

The summation part is the summation of probabilities of a Hypergeometric distribution with $N = 17$, $r = 10$, $n = 6 - y$. So it sums up to 1. The remaining part is

$$p_Y(y) = \frac{\binom{8}{y}\binom{17}{6-y}}{\binom{25}{6}}, \quad 0 \leq y \leq 6,$$

and 0 otherwise.

Solution continued
d.
Given the number of white balls in the sample as $Y = y$, the conditional pmf of X is

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{\frac{\binom{10}{x}\binom{8}{y}\binom{7}{6-x-y}}{\binom{25}{6}}}{\frac{\binom{8}{y}\binom{17}{6-y}}{\binom{25}{6}}} \\ &= \frac{\binom{10}{x}\binom{7}{6-x-y}}{\binom{17}{6-y}}, \quad 0 \le x \le 6-y, \end{aligned}$$

and 0 otherwise.

Textbook Prob. 3.17 and Extension

Let (X, Y) be a random point chosen uniformly on the region $R = \{(x, y) : |x| + |y| \leq 1\}$.

a. Sketch R. Use your sketch to find out the joint pdf.

b. Find $P(|X| + |Y| \leq 1/2)$.

c. Find the marginal densities of X and Y using your sketch. Be careful of the range of integration.

d. Find the conditional density of Y given X. Are they independent?

Solution:

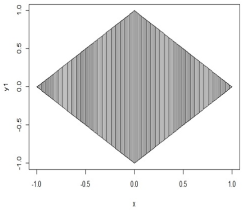


Figure: Sketch of support R.

a.
It is easy to find that the area of R is 2, so the joint pdf is

$$f_{X,Y}(x,y) = \begin{cases} 1/2, & |x| + |y| \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

Solution continued

b. There are two ways to find $P(|X| + |Y| \leq 1/2)$.

Approach 1. Obviously, this is included in R, so the probability is $\frac{1}{2}$Area of $C = \{(x, y) : |X| + |Y| \leq 1/2\}$. With similar sketch, it is easy to find the area as $1/2$, so the probability is $1/4$.

Approach 2. We use integral to solve the problem.

To use the integral method, we have to figure out the limits of integration. Figure out the limits of $X = x$ first, the limits from R is $-1 \leq x \leq 1$, and from C is $-1/2 \leq x \leq 1/2$. For fixed x, the limits for $Y = y$ is the intersection of R and C, which is $-1/2 - x$ for $x < 0$ and $x - 1/2$ for $x > 0$ as lower limits, and $1/2 + x$ for $x < 0$ and $1/2 - x$ for $x > 0$ as upper limits. So the integration can be decomposed into two parts,

$$\begin{aligned}
\int_C \int f_{X,Y}\,dy\,dx &= \int_{-1/2}^{0} \int_{-1/2-x}^{1/2+x} f_{X,Y}\,dy\,dx + \int_{0}^{1/2} \int_{x-1/2}^{1/2-x} f_{X,Y}\,dy\,dx \\
&= \int_{-1/2}^{0} \int_{-1/2-x}^{1/2+x} 1/2\,dy\,dx + \int_{0}^{1/2} \int_{x-1/2}^{1/2-x} 1/2\,dy\,dx \\
&= \int_{-1/2}^{0} (1/2 + x)\,dx + \int_{0}^{1/2} (1/2 - x)\,dx \\
&= 1/4 - 1/8 + 1/4 - 1/8 \\
&= 1/4.
\end{aligned}$$

Solution continued

c. To find the marginal density for X, we should integrate $f_{X,Y}(x,y)$ wrt $Y = y$ for all possible y. For $X = x$, the integration limits for y are $-1-x$ and $1+x$ for $x < 0$, and $x-1$ and $1-x$ for $x > 0$.

$$f_X(x) = \int_{-1-x}^{1+x} f_{X,Y}(x,y)\,dy = \int_{-1-x}^{1+x} \frac{1}{2}\,dy = 1+x, \quad -1 \le x \le 0,$$

$$f_X(x) = \int_{x-1}^{1-x} f_{X,Y}(x,y)\,dy = \int_{x-1}^{1-x} \frac{1}{2}\,dy = 1-x, \quad 0 \le x \le 1,$$

and 0 otherwise.

Similarly, the marginal pdf for Y is

$$f_Y(y) = \begin{cases} 1+y, & -1 \le y \le 0 \\ 1-y, & 0 \le y \le 1 \end{cases}$$

Solution continued

d. The conditional distribution of Y given X is achieved according to the definition $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, where

$$f_{Y|X}(y|x) = \frac{1}{2(1-|x|)}, \qquad |y| \leq 1 - |x|,$$

and 0 otherwise.

Rejection Method (Advanced Topic)

Suppose that f is a density function that is nonzero on an interval $[a, b]$ and zero outside the interval (a and b must be infinite). Let $M(x)$ be a function such that $M(x) \geq f(x)$, and let

$$m(x) = \frac{M(x)}{\int_a^b M(x)\,dx}$$

be a pdf. As we will see, the idea is to choose M so that it is easy to generate r.v.'s from m. If $f(x)$ is finite on $[a, b]$, we can choose $M(x) = \max\{f(x)\} + 1$ on the interval $[a, b]$ and 0 otherwise. Therefore, m becomes the uniform distribution on $[a, b]$. The algorithm is as follows:

Step 1. Generate T with the density m.

Step 2. Generate U, uniform on $[0, 1]$ and independent of T. If $M(T) \times U \leq f(T)$, then let $X = T$ (accept T). Otherwise, go to Step 1 (reject T).

Please check that the density function of the r.v. X obtained is f.

Rejection Method–Solution

Solution. According to the generation procedure, we generate one X if and only if we generate T and T is accepted. So,

$$
\begin{aligned}
P(x \le X \le x+dx) &= P(x \le T \le x+dx|accept) = \frac{P(x \le T \le x+dx, accept)}{P(accept)} \\
&= \frac{P(accept|x \le T \le x+dx)P(x \le T \le x+dx)}{P(accept)}
\end{aligned}
$$

First consider the numerator of this expression. We have

$$P(accept|x \le T \le x+dx) = P(U \le f(x)/M(x)) = \frac{f(x)}{M(x)}$$

so that the numerator is

$$\frac{m(x)dx\, f(x)}{M(x)} = \frac{f(x)dx}{\int_a^b M(x)dx}.$$

From the law of total probability, the denominator is

$$P(accept) = P(U \le f(T)/M(T)) = \int_a^b \frac{f(t)}{M(t)} m(t)dt.$$

Note that $m(t) = \frac{M(t)}{\int_a^b M(t)dt}$, and $\int_a^b f(t) = 1$ because f is a pdf, so

$$\int_a^b \frac{f(t)}{M(t)} m(t)dt = \int_a^b \frac{f(t)}{\int_a^b M(x)dx} dt = \frac{1}{\int_a^b M(t)dt}.$$

Finally, we see that the numerator over the denominator is $f(x)dx$.

Multiplication Law

A freshly minted coin has a certain prob. of coming up heads if it is spun on its edge, but that prob. is not necessarily equal to $1/2$. Suppose that prob. is uniformly distributed on $[0, 1]$. Now suppose it is spun n times and come up heads X times. What has been learned about the chance the coin comes up heads?

Multiplication Law–Solution

Solution. Let Θ be the probability of coming up heads for this coin, then $\Theta \sim U(0,1)$. Given θ, the conditional distribution for $X|\Theta = \theta \sim Bin(n, \theta)$. So, according to multiplication law, the joint distribution for (X, Θ) is

$$f_{X,\Theta}(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad 0 < \theta < 1, 0 \le x \le n.$$

The marginal density for $X = x$ is

$$f_X(x) = \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}d\theta = \binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}d\theta.$$

Note that the integration part is very near to Beta distribution with parameters $x+1$ and $n-x+1$. Compare with the pdf of $B(x+1, n-x+1)$ and note that the integration of pdf is 1, we have

$$\int_0^1 \theta^x(1-\theta)^{n-x}d\theta = \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n-x+1+x+1)} = \frac{x!(n-x)!}{(n+1)!}.$$

So, the marginal pmf for X becomes

$$f_X(x) = \binom{n}{x}\frac{x!(n-x)!}{(n+1)!} = \frac{n!}{x!(n-x)!}\frac{x!(n-x)!}{(n+1)!} = \frac{1}{n+1}, \quad x \in \{0, \cdots, n\}.$$

The conditional pdf for $\Theta|X = x$ is

$$f_{\Theta|X=x} = \frac{f_{X,\Theta}(x,\theta)}{f_X(x)} = (n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x} = \frac{(n+1)!}{x!(n-x)!}\theta^x(1-\theta)^{n-x}.$$

So, $[\Theta|X = x] \sim B(x+1, n-x+1)$.

Covariance and Correlation: Example 1

Suppose that X and Y have the following joint probability mass function:

	$f(x,y)$	y = 1	2	3	$f_X(x)$
x	1	.25	.25	0	.5
	2	0	.25	.25	.5
	$f_Y(y)$	.25	.5	.25	

Find the covariance and correlation coefficient between X and Y.

Solution. For this problem, the marginal distribution of X and Y is given. So we go on with Step 2.

$$E(X) = 1(.5) + 2(.5) = 1.5, \quad E(Y) = 1(.25) + 2(.5) + 3(.25) = 2.$$

Then, for Step 3, we have

$$E(XY) = (1)(1)(.25) + (1)(2)(.25) + (2)(2)(.25) + (2)(3)(.25) = 3.25.$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 3.25 - 2(1.5) = .25.$$

As we are also interested in the correlation coefficient between X and Y, so we apply Step 4, where

$$Var(X) = (1-1.5)^2(.5) + (2-1.5)^2(.5) = .25$$

$$Var(Y) = (1-2)^2(.25) + (2-2)^2(.5) + (3-2)^2(.25) = .5$$

So, with Step 5, the correlation coefficient is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{.25}{\sqrt{.25 * .5}} = 1/\sqrt{2}.$$

Covariance and Correlation: Example 2

Suppose X and Y have the joint pdf as

$$f(x,y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find $Cov(X,Y)$ and $Corr(X,Y)$.

Solution. First, we have to find the marginal pdf for X and Y. As the restriction is $0 < x < 1$, $0 < y < 1$, and $x < y$, so

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_x^1 10xy^2\,dy = 10x(1-x^3)/3, \quad 0 < x < 1,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y 10xy^2\,dx = 5y^4, \quad 0 < y < 1,$$

and 0 otherwise.
Then, we continue with Step 2,

$$E(X) = \int_0^1 x(10x(1-x^3))/3\,dx = \int_0^1 x\frac{10}{3}x^2 - \frac{10}{3}x^5\,dx = \frac{10}{9} - \frac{5}{9} = 5/9$$

$$E(Y) = \int_0^1 y5y^4\,dy = \int_0^1 5y^5\,dx = \frac{5}{6}.$$

In Step 3, we find that

$$E(XY) = \int\int xy10xy^2\,dx\,dy = \int_0^1\int_0^y 10x^2y^3 dx\,dy = \int_0^1 10y^3y^3/3\,dy = 10/21,$$

and

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{10}{21} - \frac{5}{9}\cdot\frac{5}{6} = \frac{5}{378}.$$

Solution continued
As we want to find out the correlation coefficient, we go on with Steps 4-5. In Step 4, we want to find out the variance. Note that

$$E(X^2) = \int_0^1 10x^3(1-x^3)/3\,dx = \frac{5}{6} - \frac{10}{21} = 5/14,$$

so $Var(X) = E(X^2) - (E(X))^2 = 5/14 - (5/9)^2 = \frac{55}{1134}$. Similarly,

$$E(Y^2) = \int_0^1 5y^4 \cdot y^2\,dy = \frac{5}{7},$$

so $Var(Y) = E(Y^2) - (E(Y))^2 = 5/7 - (5/6)^2 = \frac{5}{252}$.
With Step 5, the correlation coefficient is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{5/378}{\sqrt{\frac{55}{1134}\frac{5}{252}}} \approx .4264$$

Covariance and Correlation

For the above problem, what $Cov(X, 2Y)$ is? What $Cov(X + Y, 2Y)$ is? What $Corr(2X + 1, 3Y)$ is?

Solution. According to the properties of covariance and correlation coefficient,

$$Cov(X, 2Y) = 2Cov(X, Y) = 5/189.$$

$$Cov(X+Y, Y) = Cov(X, 2Y)+Cov(Y, 2Y) = 2Cov(X, Y)+2Var(Y) = 5/189+5/126 = .066$$

$$Corr(2X + 1, 3Y) = Corr(X, Y) = .4264.$$

Remark. We do not need to calculate the pdf for $2X + 3$ or $2Y$ here.

Covariance and Correlation: Remark

In example 1 and example 2, we can also calculate the mean/variance in the following way

$$E(X) = \int\int x f(x,y)\,dx\,dy, \quad Var(X) = \int\int (x - E(X))^2 f(x,y)\,dx\,dy.$$

With this method, we do not need to calculate the marginal distribution as Step 1. However, to apply the above formula, the support of $f(x,y)$ must be checked carefully.

Conditional Expectation: Example 1

There are two urns, one contains 5 red and 6 green balls, and the other one contains 3 red and 4 green balls. A player should choose one urn first randomly (without any knowledge about the balls in the run), and then choose a ball at random from the urn. If the player chooses a red ball, he/she gets a prize as \$2, otherwise he/she loses \$1. What is the expectation of the rewards for one game?

Solution. Let X be the rewards after one game for this player. Then $X = 2$ if the ball is red, and $X = -1$ if the ball is green. Let $Z = 0$ if the first urn is chosen and $Z = 1$ if the second urn is chosen, then $Z \sim Ber(1/2)$.
The conditional expectation for $X|Z$ is,

$$E[X|Z=0] = \frac{5}{5+6} \cdot 2 + \frac{6}{5+6} \cdot (-1) = \frac{4}{11}, \quad E[X|Z=1] = \frac{3}{3+4} \cdot 2 + \frac{4}{3+4} \cdot (-1) = \frac{2}{7}.$$

So, according to the law of total expectation,

$$E(X) = E[E(X|Z)] = \frac{4}{11} P(Z=0) + \frac{2}{7} P(Z=1) \approx .3247.$$

Conditional Expectation: Example 2

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel, The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours of travel. If we assume that the miner is at all times equally likely to choose any one of the doors, find the expected time for him to get to safety.

Solution. Let Y denote the door the minor initially chooses. Then we have

$$E[X|Y=1]=3, \quad E[X|Y=2]=5+E[X], \quad E[X|Y=3]=7+E[X].$$

Thus by the law of total expectation we have

$$\begin{aligned} E[X] &= E[E[X|Y]] = E[X|Y=1]P(Y=1)+E[X|Y=2]P(Y=2)+E[X|Y=3]P(Y \\ &= \frac{1}{3}(E[X|Y=1]+E[X|Y=2]+E[X|Y=3]) \\ &= 5+\frac{2}{3}E[X]. \end{aligned}$$

Therefore $E[X]=15$.

Random Sum

Suppose that the number of people entering a department store on a given day is a random variable N with mean 50. Suppose further that the amount of money spent by these customers are independent random variables with a common mean \$8. Suppose also that the amount of money spent by a customer is independent of N. Find the expected amount of money spent in the store on a given day.

Random Sum–Solution

Solution. Let X_i be the amount spent by the i-th customer. Then the total amount of money spent in the store on a given day is $\sum_{i=1}^{N} X_i$. Now

$$E(\sum_{i=1}^{N} X_i) = E\left[E(\sum_{i=1}^{N} X_i|N)\right].$$

Also,

$$E(\sum_{i=1}^{N} X_i|N=n) = E(\sum_{i=1}^{n} X_i|N=n) = E(\sum_{i=1}^{n} X_i) = nE[X_1].$$

by the independence between X_i and N.
Consequently,

$$E(\sum_{i=1}^{N} X_i|N) = NE[X_1].$$

Therefore

$$E(\sum_{i=1}^{N} X_i) = E[NE(X_1)] = E(N)E(X_1) = 400.$$

Remark. The distribution of X_i and N are not required.

Calculating moments with mgf

A company insures homes in three cities, J, K, L. The losses occurring in these cities are independent. The moment-generating functions for the loss distributions of the cities are

$$M_J(t) = (1-2t)^{-3}, \quad M_K(t) = (1-2t)^{-5/2}, \quad M_L(t) = (1-2t)^{-9/2}.$$

Let X represent the combined losses from the three cities. Calculate $E(X^3)$.

Solution. Let J, K, L denote the losses from the three cities. Then $X = J + K + L$. Since J, K, L are independent, the moment-generating function for their sum, X, is equal to the product of the individual moment-generating functions, i.e.,

$$M_X(t) = M_K(t)M_J(t)M_L(t) = (1-2t)^{-3-5/2-9/2} = (1-2t)^{-10}.$$

Differentiating the function, we get

$$\begin{aligned} M'(t) &= (-2)(-10)(1-2t)^{-11} && (4) \\ M''(t) &= (-2)^2(-10)(-11)(1-2t)^{-12} && (5) \\ M'''(t) &= (-2)^3(-10)(-11)(-12)(1-2t)^{-13} && (6) \end{aligned}$$

Hence, $E(X^3) = M_X'''(0) = (-2)^3(-10)(-11)(-12) = 10,560.$

Derivations with mgf

Suppose independent r.v.'s X and Y are distributed as $X \sim N(1,2)$, $Y \sim N(-3,4)$. What is the distribution of $Z = X + Y$?

Solution. According to the table of mgf for common distributions, the mgf for a normal r.v. with mean μ and variance σ^2 is $e^{\mu t + \sigma^2 t^2/2}$. So, the mgf for X and Y are

$$M_X(t) = e^{t+t^2}, \quad M_Y(t) = e^{-3t+2t^2}.$$

Hence,

$$M_Z(t) = M_X(t)M_Y(t) = e^{t+t^2}e^{-3t+2t^2} = e^{-2t+3t^2}.$$

Compare with the mgf for normal distribution, $Z \sim N(-2,6)$.

Convergence in Distribution (Advanced Topic)

Let $X_1, X_2, X_3, \cdots$ be a sequence of random variables such that

$$X_n \sim Geo(\lambda/n), \qquad \text{for } n = 1, 2, 3, ...$$

where $\lambda > 0$ is a constant. Define a new sequence Y_n as

$$Y_n = \frac{1}{n} X_n, \qquad \text{for } n = 1, 2, 3, ...$$

Show that Y_n converges in distribution to $exp(\lambda)$.

Solution. Note that if $W \sim Geo(p)$, then for any positive integer l, we have

$$P(W \leq l) = \sum_{k=1}^{l} (1-p)^{k-1} p = p \sum_{k=1}^{l} (1-p)^{k-1} = p \cdot \frac{1-(1-p)^l}{1-(1-p)} = 1-(1-p)^l.$$

Now, since $Y_n = \frac{1}{n} X_n$, $X_n \sim Geo(\lambda/n)$, for any positive real number, we can write

$$F_{Y_n}(y) = P(Y_n \leq y) = P(X_n \leq ny) = 1 - (1 - \frac{\lambda}{n})^{\lfloor ny \rfloor},$$

where $\lfloor ny \rfloor$ is the largest integer less than or equal to ny. We then write

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} 1 - (1 - \frac{\lambda}{n})^{\lfloor ny \rfloor} = 1 - \lim_{n\to\infty} (1 - \frac{\lambda}{n})^{\lfloor ny \rfloor} = 1 - e^{-\lambda y}.$$

The last equality holds because $ny - 1 \leq \lfloor ny \rfloor \leq ny$, and

$$\lim_{n\to\infty} (1 - \frac{\lambda}{n})^{ny} = e^{-\lambda y}.$$

Convergence in Probability

Extension of WLLN. Let $\{X_n\} \equiv X_1, \cdots, X_i, \cdots$, be a sequence of independent r.v.'s with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma_i^2$ with $\sum_{i=1}^n \sigma_i^2/n^2 \to 0$. Let $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$. Prove that $\bar{X}_n \xrightarrow{P} \mu$, where μ denotes the r.v. with probability 1 at the point μ.

- Remark 1. If $\sigma_i = \sigma$ for $i = 1, 2, 3, ...$, then it becomes the WLLN in class.
- Remark 2. If there is a universal bound for σ_i, which means that $\sigma_i \leq \sigma$ for any i, then the condition holds. It means that we do not require X_i's to share the same second moment.
- Remark 3. Even in the case that $\sigma_n^2 = \sqrt{n}$, which means that the r.v. diverges, WLLN still holds.
- Remark 4. However, if $\sigma_n^2 = n$, which means that the r.v. diverges very quickly, this WLLN does not hold.

Solution. Note that $E(\bar{X}_n) = \mu$, and

$$Var(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \to 0.$$

According to Cheybyshev Inequality,

$$P(|\bar{X}_n - \mu| \geq \epsilon) = \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sum_{i=1}^{n}\sigma_i^2}{n^2\epsilon^2} \to 0.$$

So, $\bar{X}_n \xrightarrow{P} \mu$.

Almost Sure Convergence (Advanced Topic)

Consider the sample space $S = [0, 1]$ with uniform probability distribution, i.e.,

$$P([a, b]) = b - a, \quad \text{for all } 0 \leq a \leq b \leq 1.$$

Define the sequence $\{X_n, n = 1, 2, \cdots\}$ as $X_n(s) = \frac{n}{n+1}s + (1 - s)^n$. Also, define the random variable X on this sample space as $X(s) = s$. Show that $X_n \overset{a.s.}{\to} X$.

Solution. For any $s \in (0,1]$, we have

$$\lim_{n\to\infty} X_n(s) = \lim_{n\to\infty} \left[\frac{n}{n+1}s + (1-s)^n\right] = s = X(s).$$

However, if $s = 0$, then

$$\lim_{n\to\infty} X_n(0) = \lim_{n\to\infty} \left[\frac{n}{n+1} \cdot 0 + (1-0)^n\right] = 1.$$

Thus, we conclude

$$\lim_{n\to\infty} X_n(s) = X(s), \qquad \text{for all } s \in (0,1].$$

Since $P((0,1]) = 1$, we conclude $X_n \overset{a.s.}{\to} X$.

Central Limit Theorem: Example 1

A physicist makes 25 independent measurements of the specific gravity of a certain body. He knows that the limitations of his equipment are such that the standard deviation of each measurement is σ units.

(a) By using the Chebyshev inequality, find a lower bound for the probability that the average of his measurements will differ from the actual specific gravity of the body by less than $\sigma/4$ units

(b) By using the central limit theorem, find an approximate value for the probability in part (a).

Solution.

(a) By

$$P(|\bar{X}_n - \mu| \geq \frac{\sigma}{4}) \leq \frac{\sigma^2}{n} \cdot \frac{16}{\sigma^2} = \frac{16}{25}.$$

Therefore, $P(|\bar{X}_n - \mu| \leq \frac{\sigma}{4}) \geq 1 - \frac{16}{25} = 0.36$

(b) The distribution of

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{5}{\sigma}(\bar{X}_n - \mu)$$

will be approximately a normal distribution. Therefor,

$$P(|\bar{X}_n - \mu| \leq \sigma/4) = P(|Z| \leq 5/4) \approx 2\Phi(1.25) - 1 = 0.788.$$

Central Limit Theorem: Example 2

Suppose that the distribution of the number of defects on any given bolt of cloth is the Poisson distribution with mean 5, and the number of defects on each bolt is counted for a random sample of 125 bolts. Determine the probability that the average number of defects per bolt in the sample will be less than 5.5.

Solution. Since the variance of a Poisson distribution is equal to the mean, the number of defects on any bolt has mean 5 and variance 5. Therefore, the distribution of the average number $\bar{X}_n$ on the 125 bolts will be approximately a normal distribution with mean 5 and variance $5/125 = 1/25$. If we let $Z = (\bar{X}_n - 5)/(1/5)$, then the distribution of Z will be approximately a normal distribution. Hence

$$P(\bar{X}_n < 5.5) = P(Z < 2.5) \approx \Phi(2.5) = 0.9938.$$

Central Limit Theorem: Example 3

Suppose that 75 percent of the people in a certain metropolitan area live in the city and 25 percent of the people live in the suburbs. If 1200 people attending a certain concert represent a random sample from the metropolitan area, what is the probability that the number of people from the suburbs attending the concert will be fewer than 270?

Solution. The total number of people X from the suburbs attending the concert can be regarded as the sum of 1200 independent r.v.'s, each of which has a Bernoulli distribution with parameter $p = 1/4$. Therefore, the distribution of X will be approximately a normal distribution with mean $1200(1/4) = 300$ and variance $1200(1/4)(3/4) = 225$. If we let $Z = (X - 300)/15$, then the distribution of Z will be approximately a standard normal distribution. Hence,

$$P(X < 270) = P(Z < -2) \approx 1 - \Phi(2) = 0.0228.$$

Maximum Likelihood Estimator: Example

Suppose $X \sim Unif(a, b)$, and we get n indept observation on X.

(a) Find out the MLE for a and b

(b) Find the distribution of the MLE for a and b

Maximum Likelihood Estimator–Solution

Solution:

(a) The likelihood function for $X_1, \dots X_n$ is

$$L(a,b) = \frac{1}{(b-a)^n} \prod_{i=1}^{n} I(a \le X_i \le b) = \frac{1}{(b-a)^n} I(a \le \min_i X_i) I(b \ge \max_i X_i).$$

Therefore, if $a > \min_i X_i$ or $b < \max_i X_i$, the likelihood function is 0; otherwise, $L(a,b) > 0$. So the MLE is restricted on the interval that $a \le \min_i X_i$ and $b \ge \max_i X_i$.

Note that $L(a,b)$ decreases as $b-a$ increases, so it achieves maximum when $b-a$ achieves minimum, which happens when $b = \max_i X_i$ and $a = \min_i X_i$. So the maximum likelihood estimate is

$$\hat{a} = \min_i X_i, \qquad \hat{b} = \max_i X_i.$$

(b) According to the distribution of order statistics, we have that

$$f_{\hat{a}}(x) = n \frac{1}{b-a} \frac{(b-x)^{n-1}}{(b-a)^{n-1}} = n \frac{(b-x)^{n-1}}{(b-a)^n} I(a \le x \le b)$$

$$f_{\hat{b}}(x) = n \frac{1}{b-a} \frac{(x-a)^{n-1}}{(b-a)^{n-1}} = n \frac{(x-a)^{n-1}}{(b-a)^n} I(a \le x \le b)$$