My learning notes of Math Statistics I

Yang Ge

2021 - January - 23, 12:43

Contents

1 Why

1. Why we need to learn probability theory?

   In practice, it is usually **impossible** or **undesirable** to establish a deterministic relation in observed data.

2. What can one do?

   Relax the deterministic relation by some unexplainable randomness assume the data generating mechanism (aka, model) involves randomness.

2 Moment

1. k th moment of a random variable Y: $E[Y^k] = \mu'_k$. For example, $E[Y^1] = \mu'_1 = \mu$.
2. k th central moment of Y: $E[(Y-\mu)^k] = \mu_k$. For example, $\sigma = \mu_2$.
3. Moment-generating function (Identifier for a distribution) for a random variable Y. $m(t) = E[e^{tY}]$

Example

4.136 Suppose that the waiting time for the first customer to enter a retail shop after 9:00 A.M. is a random variable Y with an exponential density function given by

$$f(y) = \begin{cases} \left(\frac{1}{\theta}\right) e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the moment-generating function for Y.

b Use the answer from part (a) to find $E(Y)$ and $V(Y)$.

a. $\because M(t) = E[e^{ty}] = \int_0^{+\infty} e^{ty}\lambda e^{-\lambda y} dy$, $\lambda = \frac{1}{\theta}$, $y > 0$

   $\therefore M(t) = \frac{1}{t\theta - 1} e^{(t-\frac{1}{\theta})y}\Big|_0^{+\infty} = \frac{1}{1-\theta t}$, if $t < \frac{1}{\theta}$

b. if $\frac{1}{\theta} > t$, $\because M'(t=0) = E[Y] = \theta$, $M''(t=0) = E[Y^2] = 2\theta^2$

   $\therefore Var(Y) = E[Y^2] - (E[Y])^2 = 2\theta^2 - \theta^2 = \theta^2$.

4. Let $Y_1, Y_2, ..., Y_n$ be independent random variables with moment-generating functions $m_{Y_1}(t), m_{Y_2}(t), ..., m_{Y_n}(t)$, respectively, if $U = Y_1 + Y_2 + ... + Y_N$, then $m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times ... \times m_{Y_n}(t)$

3 Tchebysheff's Theorem

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2} \; OR \; P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

4 Some basics

1. Let c be a constant and let $g(Y)$, $g_1(Y)$, $g_2(Y)$, ... , $g_k(Y)$ be functions of a continuous random variable Y. Then the following results hold:

- $E[c] = c$
- $E[cg(Y)] = cE[g(Y)]$

- $E[g_1(Y) + g_2(Y) + ... + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + ... + E[g_k(Y)]$

2. $\sigma^2 = V(Y) = E[Y^2] - (E[Y])^2$
3. Multivariate Probability Distributions:

- $P(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$
- $F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2)$

4. Marginal and Conditional Probability Distributions

- $P_1(y_1) = \sum\limits_{all\, y_2} P(y_1, y_2)$
- $f_1(y_1) = \int\limits_{-\infty}^{+\infty} f(y_1, y_2)\, dy_2$
- $P(y_1|y_2) = P(Y = y_1|Y_2 = y_2) = \frac{P(y_1,y_2)}{P_2(y_2)}$
- $f(y_1|y_2) = \frac{f(y_1,y_2)}{f_2(y_2)}$

Conditional expectations are defined in the same manner as univariate expectations except that conditional densities and probability functions are used in place of their marginal counterparts.

If Y_1 and Y_2 are any two random variables, the conditional expectation of $g(Y_1)$, given that $Y_2 = y_2$, is defined to be:

1. if Y_1 and Y_2 are jointly continuous:

$$E(g(Y_1)|Y_2 = y_2) = \int\limits_{-\infty}^{\infty} g(y_1)\, f(y_1|y_2)\, dy_1$$

2. if Y_1 and Y_2 are jointly discrete:

$$E(g(Y_1)|Y_2 = y_2) = \sum_{all y_1} g(y_1)\, p(y_1|y_2)$$

Properties:

1. $E[\sum x_i|Y = y] = \sum E[x_i|Y = y]$
2. $E[ax|Y = y] = aE[x|Y = y]$
3. $E[\sum a_i x_i|Y = y] = \sum a_i E[x_i|Y = y]$
4. $E[g(x)|Y = y] = \sum g(x)P(x|y)$ or $= \int g(x)f(x|y)dx$
5. Independent Random Variables

- $F(y_1, y_2) = F_1(y_1)F_2(y_2)$
- $E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)]$

6. The Covariance of Two Random Variables

- $Cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] = E(Y_1Y_2) - E(Y_1)E(Y_2)$
- If Y1 and Y2 are independent random variables $Cov(Y_1, Y_2) = 0$

7. Total expectation

$$E[X] = E\Big[E[X|Y = y]\Big]$$

1. if Y_1 and Y_2 are jointly continuous:

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y]\, f(y)\, dy_1$$

2. if Y_1 and Y_2 are jointly discrete:

$$E[X] = \sum_{ally} E[X|Y = y]\, P(y)$$

8. Functions of Random Variables

Theorem: A sequence of independent Normal random variables $(X_1, X_2, X_3, ..., X_n)$, each with μ_i and σ_i. Then $\sum X_i =$ is a normal distribution with $\mu = \sum \mu_i$, $\sigma^2 = \sum \sigma_i^2$.

Consider random variables $Y_1, Y_2, ..., Y_n$ and a function $U(Y_1, Y_2, ..., Y_n)$, denoted simply as U. Then three of the methods for finding the probability distribution of U are as follows:

Method 1:

- Write out the distribution function of U: $F_u(a) = P(u \leq a) = P\Big(U(Y_1, Y_2, ..., Y_n) \leq a\Big)$
- Find the region of $y_1, y_2, ..., y_n$ such that $U(y_1, y_2, ..., y_n) \leq a$
- Integrate $f(y_1, y_2, ..., y_n)$ over region $\int\int ... \int\limits_D f(y_1, y_2, y_3, ..., y_n) dy_1 dy_2 ... dy_n$

Method 2:

- Transformation: $f_U(u) = F_Y\Big(h^{-1}(u)\Big)\Big|\frac{d\big(h^{-1}(u)\big)}{du}\Big|$
- Joint density function: fix $X_1 = x_1$ and denote $U_{X_1=x_1} = h(X_2) = g(X_1 = x_1, X_2)$
- Calculate joint density function of X_1 and U: $f_{x_1,u}(x_1, u) = f_{x_1,x_2}(x_1, h^{-1}(u))\Big|d(\frac{h^{-1}(u)}{du}\Big|$
- Integrate joint density with X_1: $f_U(u) = \int f_{x_1,u}(x_1, u) dx_1$

9. Distribution functions:

The method of distribution functions:

```
a. Write out the distribution function of $U$: $F_u(a) = P(u \leq a) = P(U (Y_1 , Y_2 , ... , Y_n) \leq
```

```
b. Find the region of $y_1 , y_2 , . . . , y_n$ such that $U (y_1 , y_2 , ... , y_n) \leq a$. Denote as
```

```
c. Integrate $f(y_1 , y_2 , . . . , y_n)$ over region of D: $\int\int...\int\limits_D f(y_1 , y_2 , ...
```

```
d. Find $F(U)$, then $f(u) = F'(U)$
```

10. Transformations:

The method of transformations:

Let Y have probability density function $f_Y(y)$. If $h(y)$ is either increasing or decreasing for all y such that $f_Y(y) > 0$, then $U = h(Y)$ has density function:

$$f_U(u) = f_Y\left(h^{-1}(u)\right)\left|\frac{d\left(h^{-1}(u)\right)}{du}\right|$$

For calculate **Joint density function**:

$$U = g(x_1, x_2) \xrightarrow{find} f_u(u)$$

`a. Fix $X_1 = x_1$ and denote $U_{X_1 = x_1} = h(X_2) = g(X_1=x_1, X_2)$`

`b. Calculate joint density function of $X_1$ and $U$:`

$$f_{x_1,u}(x_1, u) = f_{x_1,x_2}(x_1, h^{-1}(u))\left|\frac{d\left(h^{-1}(u)\right)}{du}\right|$$

`c. Integrate joint density with $X_1$:`

$$f_U(u) = \int f_{x_1,u}(x_1, u)dx_1$$

11. Bayes' Rule:

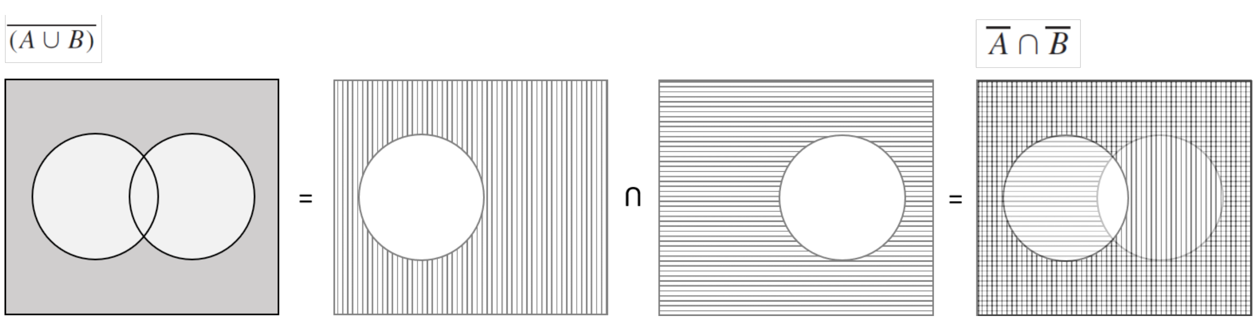
$$P(B_j|A) = \frac{P(A \bigcap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$$

12. Transform normal to standard: $Z = \frac{Y-\mu}{\sigma}$

5 Basic practices

1. Q2.3

2.3 Draw Venn diagrams to verify DeMorgan's laws. That is, for any two sets A and B, $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$ and $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$.



2. Q2.20

***2.20** The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car) and behind the other two curtains were worthless prizes (duds). The contestant was asked to choose one curtain. If the curtains are identified by their prizes, they could be labeled G, D_1, and D_2 (Good Prize, Dud1, and Dud2). Thus, the sample space for the contestants choice is $S = \{G, D_1, D_2\}$.[1]

a If the contestant has no idea which curtains hide the various prizes and selects a curtain at random, assign reasonable probabilities to the simple events and calculate the probability that the contestant selects the curtain hiding the nice prize.

b Before showing the contestant what was behind the curtain initially chosen, the game show host would open one of the curtains and show the contestant one of the duds (he could always do this because he knew the curtain hiding the good prize). He then offered the contestant the option of changing from the curtain initially selected to the other remaining unopened curtain. Which strategy maximizes the contestant's probability of winning the good prize: stay with the initial choice or switch to the other curtain? In answering the following sequence of questions, you will discover that, perhaps surprisingly, this question can be answered by considering only the sample space above and using the probabilities that you assigned to answer part (a).

i If the contestant choses to stay with her initial choice, she wins the good prize if and only if she initially chose curtain G. If she stays with her initial choice, what is the probability that she wins the good prize?

ii If the host shows her one of the duds and she switches to the other unopened curtain, what will be the result if she had initially selected G?

iii Answer the question in part (ii) if she had initially selected one of the duds.

iv If the contestant switches from her initial choice (as the result of being shown one of the duds), what is the probability that the contestant wins the good prize?

v Which strategy maximizes the contestant's probability of winning the good prize: stay with the initial choice or switch to the other curtain?

a. If the contestant has no idea which curtains hide the various prizes and selects a curtain at random, assign reasonable probabilities to the simple events and calculate the probability that the contestant selects the curtain hiding the nice prize.Because he can only choose one, so G, D1, D2 are mutually excluded, he has no idea mean they are also equally and sum up equal to 1, so Event1 = {G}, p1 = 1/3; Event2 = {Dud1}, p2 = 1/3; Event3 = {Dud2}, p3 = 1/3. Probability that the contestant selects the curtain hiding the nice prize = Event1 = {G}, p1 = 1/3.

b. If the contestant choses to stay with her initial choice, she wins the good prize if and only if she initially chose curtain G. If she stays with her initial choice, what is the probability that she wins the good prize?

$P_G = 1/3$, if she stays with her initial choice, she wins the good prize by $1/3$ chance.

ii. If the host shows her one of the duds and she switches to the other unopened curtain, what will be the result if she had initially selected G?

If she had initially selected G and switche to the other unopened curtain, then she has no chance to win, p = 0.

iii. Answer the question in part (ii) if she had initially selected one of the duds.

If she had initially selected D and switche to the other unopened curtain, then she has chance to win, p = 1.

iv. If the contestant switches from her initial choice (as the result of being shown one of the duds), what is the probability that the contestant wins the good prize?

The sample size are {(inital G, no swith), (inital G, swith), (Initial D, no swith), (Initial D, swith)} with probability are {1/3, 0, 0, x}. 1/3 + x should = 1. So, if she had initially selected one of the duds and switched, she has 2/3 chance to win.

v. Which strategy maximizes the contestant's probability of winning the good prize: stay with the initial choice or switch to the other curtain?

Switch to the other curtain maximizes the contestant's probability of winning the good prize.

2. Q2.62

2.62 A manufacturer has nine distinct motors in stock, two of which came from a particular supplier. The motors must be divided among three production lines, with three motors going to each line. If the assignment of motors to lines is random, find the probability that both motors from the particular supplier are assigned to the first line.

Sample size = {9 motors going to three production lines as groups of 3} = $\begin{pmatrix} 3 \\ 9 \end{pmatrix} * \begin{pmatrix} 3 \\ 6 \end{pmatrix} * \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 1680$.

If two from from the particular supplier are assigned to the first line, then E1 = {1th line: two + others, 2th line: three others, 3th: remain three others} = $\begin{pmatrix} 1 \\ 7 \end{pmatrix} * \begin{pmatrix} 3 \\ 6 \end{pmatrix} * \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 140$.

So, P(E1) = $\frac{140}{1680} = 0.083$.