STA2001 Probability and Statistics (I)

Lecture 15

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review of Last Lecture

Histogram for continuous distribution

The simplest form of a histogram is constructed as follows

1. divide (or "bin") the sample space of the distribution into a sequence of adjacent, non-overlapping and equally spaced subintervals.
2. treat each subinterval as an event, then count how many observed numerical outcomes fall into each subinterval and calculate the relative frequency
3. draw a rectangle erected over the bin with height equal to the relative frequency divided by the width of each subinterval.

Remark:

- Note that the area of the histogram is equal to 1, thus histogram gives an approximation of the probability density function of the underlying random variable.

Review of Last Lecture

Key concepts and/or techniques:

1. Multivariate RV, $X_1, \cdots, X_n$
2. Independence of $X_1, \cdots, X_n$
3. Random sample of size n, i.e., i.i.d.
4. If $X_1, \cdots, X_n$ are independent, then

$$E[u_1(X_1)u_2(X_2)\cdots u_n(X_n)] = E[u_1(X_1)]E[u_2(X_2)]\cdots E[u_n(X_n)]$$

5. If $X_1, \cdots, X_n$ are independent, then

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i]$$
$$Var[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i^2 Var[X_i]$$

Review of Last Lecture

- Discrete type: $X_1, X_2, \cdots, X_n$ are all discrete

  Joint pmf $f(x_1, \cdots, x_n) : \overline{S} \to (0, 1]$

  1. $f(x_1, \cdots, x_n) > 0, \quad (x_1, \cdots, x_n) \in \overline{S}$
  2. $\sum_{x_1, \cdots, x_n \in \overline{S}} f(x_1, \cdots, x_n) = 1$
  3. $P((X_1, \cdots, X_n) \in A) = \sum_{(x_1, \cdots, x_n) \in A} f(x_1, \cdots, x_n)$

- Continuous type: $X_1, X_2, \cdots, X_n$ are all continuous

  Joint pdf $f(x_1, \cdots, x_n) : \overline{S} \to (0, \infty)$

  1. $f(x_1, \cdots, x_n) > 0, \quad (x_1, \cdots, x_n) \in \overline{S}$
  2. $\int_{\overline{S}} f(x_1, \cdots, x_n) dx_1 \cdots dx_n = 1.$
  3. $P((X_1, \cdots, X_n) \in A) = \int_A f(x_1, \cdots, x_n) dx_1 \cdots dx_n.$

Review of Last Lecture

[N independent RVs]

The n RVs $X_1, \cdots, X_n$ are said to be (mutually) independent if

$$f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdot \cdots \cdot f_{X_n}(x_n),$$

where $f(x_1, \cdots, x_n)$ is the joint pmf or pdf of $X_1 \cdots X_n$, and $f_{X_i}(x_i)$ is the marginal pmf or pdf of X_i, $i = 1, \cdots, n$.

A necessary condition for the independence of the n RVs $X_1, \cdots, X_n$ is

$$\overline{S} = \overline{S_{X_1}} \times \cdots \times \overline{S_{X_n}}.$$

Remark: If $X_1, \cdots, X_n$ are independent, then any pair of them, any triple of them, $\cdots$, any $(n-1)$ of them are also independent.

Review of Last Lecture

[Theorem 5.3-1, page 191]

Assume that $X_1, X_2, \cdots, X_n$ are independent *RVs* and

$$Y = u_1(X_1)u_2(X_2)\cdots u_n(X_n)$$

If $E[u_i(X_i)], i = 1, \cdots, n$ exist. Then

$$\begin{aligned} E[Y] &= E[u_1(X_1)u_2(X_2)\cdots u_n(X_n)] \\ &= E[u_1(X_1)]E[u_2(X_2)]\cdots E[u_n(X_n)] \end{aligned}$$

Remark: This is an extension of the result that when X and Y are independent, $E(XY) = E(X)E(Y)$.

Review of Last Lecture

[Theorem 5.3-2, page 192]

Assume that $X_1, X_2, \cdots, X_n$ are independent RVs with respective mean $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, respectively. Consider $Y = \sum_{i=1}^n a_i X_i$, where $a_1, a_2, \cdots, a_n$ are real constants. Then

$$E(Y) = \sum_{i=1}^n a_i \mu_i \quad \text{and} \quad Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

Statistic

Definition

Any function of the random sample $X_1, X_2, \cdots, X_n$ that do not have any unknown parameters is called a statistic.

Definition

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed with mean μ. Then the sample mean is defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and a <u>statistic</u> and also an <u>estimator</u> of mean μ.

In what follows, we will study the properties of the sample mean $\overline{X}$.

Section 5.4 Moment generating function technique

Motivation and Goal

Motivation

Mgf, if exists, uniquely determines the distribution of the RV. Therefore, the distribution of a RV can be equivalently found via its mgf.

Goal

To derive the distribution of functions of multivariate RVs $X_1, \cdots, X_n$ with mgf technique, where the function takes the form of

$$Y = \sum_{i=1}^{n} a_i X_i$$

Example 1, page 195

Example

Let X_1 and X_2 be independent RVs with uniform distribution on $\{1, 2, 3, 4\}$. Let $Y = X_1 + X_2$. What is the distribution of Y, i.e., pmf of Y?

Example 1, page 195

Example

Let X_1 and X_2 be independent RVs with uniform distribution on $\{1, 2, 3, 4\}$. Let $Y = X_1 + X_2$. What is the distribution of Y, i.e., pmf of Y?

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E[e^{t(X_1+X_2)}] \\ &= E(e^{tX_1}) \cdot E(e^{tX_2}) \\ &= M_{X_1}(t) \cdot M_{X_2}(t) \end{aligned}$$

Example 1, page 195

Example

Let X_1 and X_2 be independent RVs with uniform distribution on $\{1, 2, 3, 4\}$. Let $Y = X_1 + X_2$. What is the distribution of Y, i.e., pmf of Y?

$$\begin{aligned} M_Y(t) = E(e^{tY}) &= E[e^{t(X_1+X_2)}] \\ &= E(e^{tX_1}) \cdot E(e^{tX_2}) \\ &= M_{X_1}(t) \cdot M_{X_2}(t) \end{aligned}$$

$$f(x) = \frac{1}{4}, \quad x = 1, 2, 3, 4$$

$$\Rightarrow M_X(t) = E(e^{tX}) = \sum_{x=1}^{4} f(x)e^{tx} = \frac{1}{4}\sum_{x=1}^{4} e^{tx}$$

Example 1, page 195

$$
\begin{aligned}
M_Y(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \\
&= \left(\frac{1}{4} \sum_{x_1=1}^{4} e^{tx_1} \right) \left(\frac{1}{4} \sum_{x_2=1}^{4} e^{tx_2} \right) \\
&= \frac{1}{16} e^{2t} + \frac{2}{16} e^{3t} + \frac{3}{16} e^{4t} + \frac{4}{16} e^{5t} + \frac{3}{16} e^{6t} + \frac{2}{16} e^{7t} + \frac{1}{16} e^{8t}
\end{aligned}
$$

Example 1, page 195

$$
\begin{aligned}
M_Y(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \\
&= \left(\frac{1}{4}\sum_{x_1=1}^{4} e^{tx_1}\right)\left(\frac{1}{4}\sum_{x_2=1}^{4} e^{tx_2}\right) \\
&= \frac{1}{16}e^{2t} + \frac{2}{16}e^{3t} + \frac{3}{16}e^{4t} + \frac{4}{16}e^{5t} + \frac{3}{16}e^{6t} + \frac{2}{16}e^{7t} + \frac{1}{16}e^{8t}
\end{aligned}
$$

Then the pmf of Y can be derived as follows

$$
\begin{aligned}
\overline{S_Y} &= \{2, 3, \cdots, 8\} \\
g(y) &= P(Y = y) = \text{the coefficient of } e^{yt}, y \in \overline{S_Y}.
\end{aligned}
$$

Theorem 5.4-1, page 196

If $X_1, X_2, \cdots, X_n$ are independent RVs with respective mgfs $M_{X_i}(t)$ where $|t| < h_i$ for $h_i > 0, i = 1, 2, \cdots, n$. Then the mgf of $Y = \sum_{i=1}^{n} a_i X_i$ is

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t),$$

where $|a_i t| < h_i, i = 1, \cdots, n$.

Proof of Theorem 5.4-1, page 196

$$M_Y(t) = E[e^{tY}] = E[e^{t\sum_{i=1}^{n} a_i X_i}]$$

$$= E[e^{ta_1X_1} e^{ta_2X_2} \cdots e^{ta_nX_n}]$$

$$\xlongequal{\text{by Thm5.3-1 page 191}} E[e^{a_1tX_1}] \cdots E[e^{a_ntX_n}]$$

$$\xlongequal{M_X(t)=E[e^{tX}]} \prod_{i=1}^{n} M_{X_i}(a_i t)$$

$M_{X_i}(t)$ is defined for $|t| < h_i$,
$M_{X_i}(a_i t)$ is defined for $|a_i t| < h_i$.

Corollary 5.4-1, page 197

If $X_1, X_2, \cdots, X_n$ is a random sample of size n from a distribution with mgf $M(t)$, where $|t| < h$, then

(a) The mgf of $Y = \sum_{i=1}^{n} X_i$, is

$$M_Y(t) = \prod_{i=1}^{n} M(t) = (M(t))^n, \quad |t| < h$$

(b) The mgf of $\overline{X} = \sum_{i=1}^{n} \frac{1}{n} X_i$ is

$$M_{\overline{X}}(t) = \prod_{i=1}^{n} M\left(\frac{1}{n}t\right) = \left[M\left(\frac{t}{n}\right)\right]^n, \quad \left|\frac{t}{n}\right| < h$$

Example 2, page 197

Let $X_1, X_2, \cdots, X_n$ denote the outcome of n Bernoulli trials each with probability of success p. Let $Y = \sum_{i=1}^{n} X_i$, then what is the distribution of Y?

Example 2, page 197

Let $X_1, X_2, \cdots, X_n$ denote the outcome of n Bernoulli trials each with probability of success p. Let $Y = \sum_{i=1}^{n} X_i$, then what is the distribution of Y?

Recall that the mgf of $X_i, i = 1, 2, \cdots, n$ is

$$M(t) = 1 - p + pe^t, \quad -\infty < t < \infty$$

Then by Corollary 5.4-1,

$$M_Y(t) = \prod_{i=1}^{n}(1 - p + pe^t) = (1 - p + pe^t)^n \Longrightarrow Y \sim b(n, p)$$

Theorem 5.4-2, page 198

Let $X_1, X_2, \cdots, X_n$ be independent chi-square RVs with

$r_1, r_2, \cdots, r_n$ degrees of freedom, respectively, i.e.,

$X_i \sim \chi^2(r_i), i = 1, \cdots, n$. Then

$$Y = X_1 + X_2 + \cdots + X_n \quad \text{is} \quad \chi^2(r_1 + r_2 + \cdots + r_n)$$

Proof of Theorem 5.4-2, page 198

Recall from Thm 5.4-1, if $X_1, \cdots, X_n$ are independent RVs

with respective mgfs $M_{X_i}(t), |t| < h_i, i = 1, \cdots, n$. Then

$Y = \sum_{i=1}^{n} X_i$ has the mgf

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t), |t| < h_i, i = 1, \cdots, n.$$

Then recall that the mgf for chi-square distribution with degree of

freedom r_i is

$$M_{X_i}(t) = (1 - 2t)^{-\frac{r_i}{2}}$$

Proof of Theorem 5.4-2, page 198

Then we have

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$= (1-2t)^{-\frac{r_1}{2}}(1-2t)^{-\frac{r_2}{2}} \cdots (1-2t)^{-\frac{r_n}{2}}$$

$$= (1-2t)^{-\frac{1}{2}(r_1+\cdots+r_n)} \Rightarrow Y \sim \chi^2(r_1+\cdots+r_n)$$

Corollary 5.4-2, page 198

Let $Z_1, Z_2, \cdots, Z_n$ have standard normal distributions, $N(0, 1)$. If these random variables are independent, then

$$W = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2(n)$$

Proof of Corollary 5.4-2, page 198

Recall from Theorem 3.3-2, if

$$X \sim N(\mu, \sigma^2) \quad \text{with} \quad \sigma^2 > 0,$$

then

$$\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$$

For the current case, $Z_i^2 \sim \chi^2(1)$, $i = 1, \cdots, n$. Then by Theorem 5.4-2 and the independence of $Z_1, \cdots, Z_n$,

$$W = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n).$$

Corollary 5.4-3, page 198

If $X_1, X_2, \cdots, X_n$ are independent and have normal distributions $N(\mu_i, \sigma_i^2), i = 1, 2, ..., n$, respectively, then the distribution of

$$W = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(n)$$

Proof: Let $Z_i = \frac{X_i - \mu_i}{\sigma_i}$, obviously, $Z_i \sim N(0, 1)$ and $Z_i^2 \sim \chi^2(n)$, $i = 1, \cdots, n$. Then by Theorem 5.4.2 and the independence of $X_1, \cdots, X_n$, we complete the proof.

Section 5.5 Random function associated with normal distribution

Theorem 5.5-1, page 200

[Theorem 5.5-1]

If $X_1, X_2, \cdots, X_n$ are n independent normal variables with means $\mu_1, \mu_2, \cdots, \mu_n$ and variances σ_1^2, σ_2^2, $\cdots$, σ_n^2, respectively, then $Y = \sum_{i=1}^{n} a_i X_i$ has the normal distribution

$$Y \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Proof of Theorem 5.5-1, page 200

By Theorem 5.4-1, we have

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t) = \prod_{i=1}^{n} \exp(\mu_i a_i t + \frac{1}{2}\sigma_i^2 a_i^2 t^2)$$

$$= \exp\left\{ \left(\sum_{i=1}^{n} \mu_i a_i \right) t + \frac{1}{2} \left(\sum_{i=1}^{n} a_i^2 \sigma_i^2 \right) t^2 \right\}$$

Example 1, page 201

Let X_1 and X_2 be the pounds of butter fat produced by 2 cows, respectively. Assume that

$$X_1 \sim N(693.2, 22820), X_2 \sim N(631.7, 19205)$$

and moreover, X_1 and X_2 are independent. What's the probability $P(X_1 > X_2)$?

Example 1, page 201

Let $Y = X_1 - X_2$. Then

$$Y \sim N(693.2 - 631.7, 22820 + 19205) = N(61.5, 42025)$$

$$P(X_1 > X_2) = P(Y > 0) = P\left(\frac{Y - 61.5}{\sqrt{42025}} > \frac{0 - 61.5}{\sqrt{42025}}\right)$$

$$= 1 - \Phi(-0.3) = 0.6179.$$

Corollary 5.5-1, Page 201

[Corollary 5.5-1]

If $X_1, X_2, \cdots, X_n$ is a random sample of size n from the normal distribution $N(\mu, \sigma^2)$, then the sample mean $\overline{X}$ has the following distribution

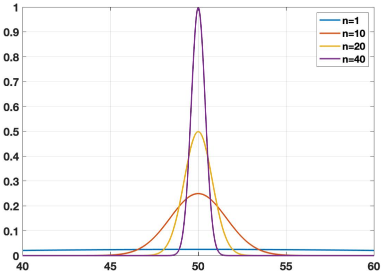
$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Leftrightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Proof: Let $a_i = \frac{1}{n}, \mu_i = \mu, \sigma_i^2 = \sigma^2, i = 1, \cdots, n$. Then by Theorem 5.5-1, we obtain the result.

Example 2, page 201

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from $N(50, 16)$,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(50, \frac{16}{n})$$



- To illustrate the effect of n: The larger n, the smaller the variance $\frac{16}{n}$.
- pdf of $\overline{X}$: The sharper the peak, the more concentrated in a small interval centered at 50.

Sample Variance

Definition

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed with mean μ and variance σ^2. Then the sample variance is defined as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2,$$

and an estimator of the variance σ^2, because

$$E(S^2) = \sigma^2.$$

Sample Variance

Note that

$$\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}\left(X_i-\overline{X}+\overline{X}-\mu\right)^2$$

$$=\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2+\sum_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu)+\sum_{i=1}^{n}\left(\overline{X}-\mu\right)^2$$

where

$$\sum_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu)=(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\overline{X})$$

$$=(\overline{X}-\mu)(\sum_{i=1}^{n}X_i-n\overline{X})=(\overline{X}-\mu)(n\overline{X}-n\overline{X})=0$$

Therefore,

$$S^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2-\sum_{i=1}^{n}\left(\overline{X}-\mu\right)^2\right]$$

Sample Variance

Then note that

$$E\left(\sum_{i=1}^{n}(X_i-\mu)^2\right)=\sum_{i=1}^{n}E\left[(X_i-\mu)^2\right]=n\cdot\sigma^2$$
$$E\left(\sum_{i=1}^{n}\left(\overline{X}-\mu\right)^2\right)=\sum_{i=1}^{n}E\left[\left(\overline{X}-\mu\right)^2\right]=n\cdot\frac{\sigma^2}{n}=\sigma^2$$

Therefore,

$$E(S^2)=\frac{1}{n-1}\left(n\sigma^2-\sigma^2\right)=\sigma^2$$