STATS 412

Sixth Class Note

In Son Zeng

29 September, 2018

My Office Hour:

My office hours are on **16:30 - 18:00 Tuesday** and **13:30 - 15:00 Friday**, at **USB 2165**. You may check the campus map to get to my office.

Calculus Review:

The calculus review will be held at USB 2165, inside the study room, at **14:00 - 15:30 Sunday**. The topics include:

• U-substitution and integration by parts for single variable integration

• Double integration, Assigning bounds for iterative integral

Reminders for Assignment 3

• Please correctly identify the CDF $F(X)$, the cumulative distribution function. Let X be a random variable defined on $[a, b]$, then the (integral/summation of discrete points) is only valid for x inside the domain (support); the cumulative distribution function is $F(x) = 0, x < a$ and $F(x) = 1, x \geq b$.

• To derive the standard deviation for the summation of independent random variables, say $X_1+X_2+......+X_n$, **please compute the variance first** by formula $Var(X_1 + X_2 + ... + X_n) = Var(X_1) + ... + Var(X_n)$. Then, you can take the square root to obtain the standard deviation. **Do not directly add the standard deviations.**

• The notion that we repeat the same trial or experiment until obtaining the first success/failure/(something) indicates that the random variable X follows geometric distribution. It is a discrete probability distribution with probability mass function:

$$\boxed{P(X = k) = \begin{cases} (1-p)^{k-1} \cdot p, k = 1, 2, 3, ... \\ 0, \ otherwise \end{cases}} \tag{1}$$

Here splitting the case is important; it distinguishes the points which we have the probability value different from 0.

• We can extend that notion to repeat the same trial or experiment until the r success/failure occurs. In this case, the random variable X follows negative binomial distribution. It is also a discrete probability distribution, with probability mass function:

$$\boxed{P(X = n) = \begin{cases} \binom{n-1}{r-1}(1-p)^{n-r}p^{r-1} \cdot p = \binom{n-1}{r-1}(1-p)^{n-r}p^r, n = r, r+1, r+2, ... \\ 0, \ otherwise \end{cases}} \tag{2}$$

• If $X_1, X_2, ..., X_n$ is a simple random sample from a population with mean μ and variance σ^2, then the sample mean $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ is a random variable with

$$\boxed{\mu_{\bar{X}} = E(\bar{X}) = E\Big(\frac{\sum_{i=1}^n X_i}{n}\Big) = \frac{E(X_1) + E(X_2) + ... + E(X_n)}{n} = \frac{n \cdot \mu}{n} = \mu} \tag{3}$$

$$\boxed{\sigma^2_{\bar{X}} = Var(\bar{X}) = Var\Big(\frac{\sum_{i=1}^n X_i}{n}\Big) = \frac{Var(X_1) + Var(X_2) + ... + Var(X_n)}{n^2} = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}} \tag{4}$$

$$\boxed{\sigma_{\bar{X}} = SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}} \tag{5}$$

Key Points during Lecture:

Conditional Probability Clarification: First of all, conditional probability should be a proper probability distribution. That is, for discrete random variable X, Y, the conditional probability should sum up to 1.

Then we should be aware about the notational equivalence:

•$P(y|2) = P(y|x = 2) = \frac{p(x=2,y)}{p_x(2)}$

•$P(y \leq 2|x = 2) = P(y \leq 2|2) = P(y = 2|2) + P(y = 1|2) = P(2|2) + P(1|2)$

It is encouraged that you know about the notational equivalences. However, when you are writing the homework, it is better to make the notation as clear as possible.

Conditional Probability: During the class there is a great question about conditional probability which is not intuitively easy:

$$\boxed{P(Y > 2000|X > 1500) = \frac{P(X > 1500 \cap Y > 2000)}{P(X > 1500)} = \frac{\int\int f(x,y)dydx}{\int_{1500}^{\infty} f_x(x)dx}} \tag{6}$$

First of all, recall that $f(x,y) = 0.000006e^{-0.001x-0.002y}$ for $0 < x < y < \infty$ (0 otherwise).

Then, $f_X(x) = \int_x^{\infty} 0.000006e^{-0.001x-0.002y}dy = 0.003e^{-0.001x} \cdot [-e^{-0.002y}]_x^{\infty} = 0.003e^{-0.003x}, x > 0$ (0 otherwise).

Be careful here, we do not want to integrate Y first but instead, doing this:

$$\boxed{P(Y > 2000|X > 1500) = \frac{\int_{2000}^{\infty}\int_{1500}^{y} 0.000006e^{-0.001x-0.002y}dxdy}{\int_{1500}^{\infty} 0.003e^{-0.003x}dx} = \frac{\int_{2000}^{\infty} e^{-0.002y}[-0.006e^{-0.001x}]_{1500}^{y}dy}{[-e^{-0.003}]_{1500}^{\infty}}} \tag{7}$$

$$\boxed{= \frac{\int_{2000}^{\infty}\left(0.006e^{-0.002y}e^{-1.5} - 0.006e^{-0.003y}\right)dy}{e^{-4.5}} = e^{4.5}\Big\{[-3e^{-0.002y}]_{2000}^{\infty}e^{-1.5} + [2e^{-0.003y}]_{2000}^{\infty}\Big\}} \tag{8}$$

$$\boxed{= e^{4.5}[-2e^{-6} + 3e^{-4}e^{-1.5}] = 3e^{-1} - 2e^{-1.5} \approx 0.6574} \tag{9}$$

Definition of Independence: If random variables $X_1, X_2, ..., X_n$ are independent, then (1) If $X_1, X_2, ..., X_n$ are jointly discrete, the joint pmf is equal to the product of the marginals:

$$\boxed{p(x_1, x_2, ..., x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot ...... \cdot p_{X_n}(x_n)} \tag{10}$$

(2) If $X_1, X_2, ..., X_n$ are jointly continuous, the joint pdf is equal to the product of the marginals:

$$\boxed{f(x_1, x_2, ..., x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot ...... \cdot f_{X_n}(x_n)} \tag{11}$$

(3) If X, Y are independent and jointly discrete, and given respectively the marginal of x $p_X(x) > 0$ and marginal of y $p_Y(y) > 0$, then

$$\boxed{p(x,y) = p_X(x) \cdot p_Y(y) = p_{Y|X}(y|x) \cdot p_X(x) = p_{X|Y}(x|y) \cdot p_Y(y)} \tag{12}$$

Therefore, making comparison of terms, we have $p_{Y|X}(y|x) = p_Y(y)$ and $p_{X|Y}(x|y) = p_X(x)$.

(4) If X, Y are independent and jointly continuous, and given respectively the marginal of x $f_X(x) > 0$ and marginal of y $f_Y(y) > 0$, then

$$\boxed{f(x,y) = f_X(x) \cdot f_Y(y) = f_{Y|X}(y|x) \cdot f_X(x) = f_{X|Y}(x|y) \cdot f_Y(y)} \tag{13}$$

Therefore, making comparison of terms, we have $f_{Y|X}(y|x) = f_Y(y)$ and $f_{X|Y}(x|y) = f_X(x)$.

Check Independence: For discrete random variables X and Y, to check independence, we basically need to check every pair of $p(x) \cdot p(y) = p(x,y)$. There are tricky cases when the result holds true for some pairs but not all pairs. Next week I believe you will encounter the concept of covariance; it is good and important for your preparation that generally, $cov(X,Y) = 0 \nrightarrow X, Y$ *independent*.

Last Comment:

Please inform me to fix the typos and grammatical mistakes if they exist. It is a great practice of writing and I appreciate your help!