STATS 412

Fourth Class Note

In Son Zeng

19 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.

Comments for Assignment 1 and 2

See another attached document.

Key points during lecture:

Expectation, Variance and Standard Deviation: To recap, for **discrete** random variable X, the expectation, variance and standard deviation are given by:

$$\boxed{E(X) = \mu_x = \sum_x x \cdot P(X = x)} \tag{1}$$

$$\boxed{Var(X) = \sigma_x^2 = \sum_x (x - \mu_x)^2 \cdot P(X = x) = E(X^2) - [E(X)]^2 = \sum_x x^2 \cdot P(X = x) - \Big(\sum_x x \cdot P(X = x)\Big)^2} \tag{2}$$

$$\boxed{SD(X) = \sqrt{Var(X)} = \sigma_x} \tag{3}$$

To recap, for **continuous** random variable X, the expectation, variance and standard deviation are given by:

$$\boxed{E(X) = \mu_x = \int_{-\infty}^{\infty} x \cdot f(x)dx} \tag{4}$$

$$\boxed{Var(X) = \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 \cdot f(x)dx = E(X^2) - [E(X)]^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \Big(\int_{-\infty}^{\infty} x \cdot f(x)dx\Big)^2} \tag{5}$$

$$\boxed{SD(X) = \sqrt{Var(X)} = \sigma_x} \tag{6}$$

The expectation for **continuous** distribution could be derived by integrating the products of x and the probability density function $f(x)$ over the domain. In most of the cases, the domain should be $-\infty < x < \infty$, $0 < x < \infty$ or $a \leq x \leq b$. Reminder: don't try to round the expectation to integer.

After obtaining the expectation, there are two major ways to compute the variance: (1) to integrate over the products of $[x - E(X)]^2$ and the probability density function $f(x)$ over the domain or (2) integrate over the products of x^2 times the probability density function $f(x)$ and subtract the result by the square of the expectation.

Finally, to obtain the standard deviation, we can take the positive square root of variance. Reminder: standard deviation must be positive. **Please ask questions during the office hours if you are confused between discrete and continuous random variables**.

Test reminder: If the notation $f(x)$ appears on test questions, it means that X is a **continuous random variable**. On the other hand, if the notation $p(x)$ appears on test questions, it means that X is a **discrete random variable**.

Complements: Sometimes it is more convenient to calculate the complement than computing and adding up every part of the probability. This is particularly evident when you encounter the key words like **"at least one"** and **"not all"**.

Second Central Moment (Optional): Method of moment is one of the statistics inference method to give an estimate for continuous random variable X. Now we know $E(X) = \mu$, we define the n-th central moment of a continuous random variable as:

$$\boxed{\mu_n = E\Big[(X - E(X))^n\Big] = \int_{-\infty}^{\infty} (x - \mu)^n \cdot f(x)dx} \tag{7}$$

Let us derive the 0^{th}, first and second central moments:

- Zero: $\mu_0 = E\Big[(X - E(X))^0\Big] = \int_{-\infty}^{\infty} 1 \cdot f(x)dx = 1$, obvious?
- First: $\mu_1 = E\Big[(X - E(X))^1\Big] = E\Big[X - E(X)\Big] = E(X) - E(X) = 0$, great?
- Second: $\mu_2 = E\Big[(X - E(X))^2\Big] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx = Var(X)$. This tells that variance is the second central moment of a continuous random variable X.

Gamma distribution and its Family (Optional):

Gamma distribution is crucial in probability theories since it can model disease, predict weather, and model waiting time in queue model. Gamma distribution takes a random variable X and two parameters α (shape parameter) and β (scale parameter).

The density of gamma(α, β) distribution is given by:

$$\boxed{f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}, \ \alpha > 0, \ \beta > 0, \ 0 < x < \infty} \tag{8}$$

Then, its expectation and variance are derived by:

$$\boxed{E(X) = \int_0^\infty x \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx = \int_0^\infty \frac{x^{(\alpha+1)-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx = \frac{\Gamma(\alpha+1)}{\beta^\alpha \Gamma(\alpha) (\frac{1}{\beta})^{\alpha+1}} = \alpha \cdot \beta} \tag{9}$$

$$\boxed{E(X^2) = \int_0^\infty x^2 \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx = \int_0^\infty \frac{x^{(\alpha+2)-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx = \frac{\Gamma(\alpha+2)}{\beta^\alpha \Gamma(\alpha) (\frac{1}{\beta})^{\alpha+2}} = (\alpha+1)\alpha \cdot \beta^2} \tag{10}$$

Therefore, the variance of such gamma distribution is

$$\boxed{V(X) = E(X^2) - [E(X)]^2 = (\alpha+1)\alpha \cdot \beta^2 - \alpha^2 \cdot \beta^2 = \alpha \cdot \beta^2} \tag{11}$$

To recap, the expectation of Gamma distribution is the product of the shape and scale parameter, while the variance of Gamma distribution is the product of the shape parameter and the square of the scale parameter.

Question: Could any of you try to derive the expected amount of snow on a day that it snows by using the Gamma distribution? This may serve as a more convenient way to compute than doing the integration by part!

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!