VE401 SU2022 RC week3

Shuyu Wu

UM-SJTU Joint Institute

JOINT INSTITUTE
交大密西根学院

wushuyu2002@sjtu.edu.cn

May 27, 2022

Outline

Continuous Random Variables

Probability Density Function (for non-discrete random variable): For a non-discrete random variable X, if there exists a function $f_X(x)$ such that

$$P[a \le X \le b] = \int_a^b f_X(x)dx$$

and

$$f_X(x) \ge 0, \int_{-\infty}^{+\infty} f_X(x)dx = 1$$

If X have PDF defined above, then we say X is a continuous random variable.

Cumulative Distribution Function(CDF)

The **cumulative distribution function** of a random variable is defined as $P[X \leq x]$. It's monotonous without decreasing, and can be denoted as $F_X(x)$.
If $F_X(x)$ is differentiable, then the derivative is the PDF($F'_X(x) = f_X(x)$), and X is a continuous random variable.

Expectation and Variance

The expectation of a continuous random variable is defined as follow:

$$E[X] := \int_{-\infty}^{+\infty} x f_X(x) dx$$

If the integral above converges absolutely.
The variance is defined as:

$$Var[X] := E[X^2] - E[X]^2$$

And for a given function φ, we have

$$E[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx$$

Outline

Exponential Distribution

A continuous random variable with PDF $f_X(x) = \beta e^{-\beta x} (x \geq 0)$ is called an exponential distribution with parameter β.

Some important properties of exponential distribution:

- $E[X] = \frac{1}{\beta}$
- $Var[X] = \frac{1}{\beta^2}$
- m.g.f: $m_X(t) = (1 - \frac{t}{\beta})^{-1}$

Memoryless Property

The exponential distribution has a special property called "memoryless property", i.e.,

$$P[X > s + t | X > s] = P[X > t]$$

And vice versa. It's the only continuous distribution that satisfy this property.
An LED's lifetime is likely to follow an exponential distribution. That is to say, when we have already used an LED for s hours and it's not broken, the expected lifetime is the same as a new LED. The LED won't "remember" that it has already used for s hours.
Human lifetime doesn't follow an exponential distribution. It's not memoryless.

Gamma Distribution

Has two parameter: α and β. It's PDF is tedious and you do not need to remember it. When $\alpha = 1$, this is the exponential distribution.

If X follows a Gamma distribution with parameter α and β,

- $E[X] = \frac{\alpha}{\beta}$
- $Var[X] = \frac{\alpha}{\beta^2}$
- m.g.f $m_X(t) = (1 - \frac{t}{\beta})^{-\alpha}$

The sum of α i.i.d random variable with parameter β is the Gamma distribution with parameter α and β. This can be seen from the m.g.f.

Connection With Poission Distribution

In a Poission process, the probability to get x arrivals is $\frac{(\lambda t)^x}{x!}e^{-\lambda t}$. It's a Poission distribution with parameter λt.

The time to the first arrival T follows an exponential distribution with parameter λ.

The time to the r^{th} arrival T_r follows a Gamma distribution with parameter r and λ.

Outline

Normal Distribution

The PDF of **normal distribution** is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}$$

The random variable with such PDF is said to follow a normal distriution with parameter μ and σ, denoted as $X \sim N(\mu, \sigma)$(in Horst's slide) or $X \sim N(\mu, \sigma^2)$(in other textbook).
The expectation is μ and the variance is σ^2. The standard deviation is σ.
The m.g.f of normal distribution is $e^{\mu t + \frac{\mu^2 t^2}{2}}$. Again, from the m.g.f we can see that

$$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2) \Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Standard Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $Z := \frac{X-\mu}{\sigma} \sim N(0,1)$ is said to follow a standard normal distribution.
The PDF for standard normal distribution is $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
The CDF of standard normal distribution is denoted as $\Phi(z)$.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

Unluckily, it's not an elementary function. We look up a table of Φ value in practice. In python, we can use 'scipy.stats.norm.cdf(x)' to get the value of $\Phi(x)$.

ex 3.1

1. The mass of a kind of product follows a normal distribution with mean 200 and variance 100. Randomly pick a product, what's the probability that the mass is in the range of [185,210]?

2. If the mass of the product is within the top 5%, we regard them as "heavier than normal". What's the threshold for a product to be "heavier than normal"?

ex 3.1 answer

1.

$$
\begin{aligned}
&\sigma = 10 \\
&P[185 \leq X \leq 210] \\
=&P[\frac{185-200}{10} \leq \frac{X-200}{10} \leq \frac{210-200}{10}] \\
=&P[-1.5 \leq Z \leq 1] \\
=&\Phi(1) - \Phi(-1.5) \\
=&\Phi(1) - (1 - \Phi(1.5)) = 0.8413 - (1 - 0.9332) \\
=&0.7745
\end{aligned}
$$

ex 3.1 answer

2.

$$
\begin{aligned}
&P[X \le x] = 0.95 \\
&P[Z \le \frac{x-200}{10}] = 0.95 \\
&\Phi(\frac{x-200}{10}) = 0.95 \\
&\frac{x-200}{10} = 1.64 \\
&x = 216.4
\end{aligned}
$$

3σ rule

3σ rule: The sample of a normal random variable is very likely to fall in the range of $\mu \pm 3\sigma$. The probability is $\Phi(3) - (1 - \Phi(3)) = 0.997$.
On the other hand, if you randomly pick a sample and find that it falls out of that range, you have evidence to suspect the correctness of μ or σ. This is the topic of statistics.

The Chebyshev Inequality

For a random variable X. If its expectation and variance exist, then

$$P[-m\sigma < X - \mu < m\sigma] \geq 1 - \frac{1}{m^2}$$

Since X can be any random variable, this estimation is very rough. However, it plays an important role in theoretical proof, like the proof of the weak law of large number.

Transformation of Random Variables

For a continuous random variable X, we know it's PDF $f_X(x)$. Now we want to calculate the PDF of Y, where $Y = \varphi(X)$.
General method: From $f_X(x)$ we can derive X's CDF $F_X(x)$, which is $P[X \le x]$. On the other hand, the CDF of Y, $P[Y \le y] = P[\varphi(X) \le y]$. We can then solve this inequality and interpret these intervals with the CDF of X.
If φ is monotonically increasing, then $P[Y \le y] = P[X \le \varphi^{-1}(y)] = F_X(\varphi^{-1}(y))$. Take derivative and we get $f_Y(y) = f_X(\varphi^{-1}(y))\frac{d\varphi^{-1}(y)}{dy}$. The proof is same if φ is monotonically decreasing. The result is $f_Y(y) = f_X(\varphi^{-1}(y))|\frac{d\varphi^{-1}(y)}{dy}|$.
When φ is not monotonic we don't have a formula, and you need to do the process above.

ex 3.2

Calculate the PDF of $Y := Z^2$, where Z is the standard normal distribution.
$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
Can you use the formula $f_Y(y) = f_X(\varphi^{-1}(y))|\frac{d\varphi^{-1}(y)}{dy}|$?

ex 3.2 answer

You can't use that formula because $\varphi(Z) = Z^2$ isn't a monotonic function. You need to use the general method.

$$
\begin{aligned}
&P[Y \leq y] \\
=&P[Z^2 \leq y] \\
=&P[-\sqrt{Y} \leq Z \leq \sqrt{Y}] \\
=&P[Z \leq \sqrt{Y}] - P[Z \leq -\sqrt{Y}] \\
=&\Phi(\sqrt{y}) - (1 - \Phi(\sqrt{y})) = 2\Phi(\sqrt{y}) - 1
\end{aligned}
$$

Take derivative and we get:

$$f_Y(y) = 2f_Z(\sqrt{y}) * \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}$$

This is a **chi-squared distribution** with freedom 1.

Outline

Central Limit Theorem

There're 3 versions of Central Limit Theorem:

1. de Moivre - Laplace: Approximate $B(n, p)$ to normal distribution
2. Lindeberg-Levy: Approximate n i.i.d random variable's sum to normal distribution
3. Lyapunov: Approximate n independent random variable's sum to normal distribution(no need to be identical distributed, but with some other condition)

Central Limit Theorem(de Moivre - Laplace)

When n is very large, the PDF of $B(n,p)$ is approximately $N(np, np(1-p))$. i.e.,

$$\lim_{n\to\infty} P[a < \frac{S_n - np}{\sqrt{np(1-p)}} < b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

Half-Unit Correction: $P[X \le y] \approx \Phi(\frac{y+0.5-np}{\sqrt{np(1-p)}})(y \in \mathbb{Z})$

ex 3.3

$X \sim B(100, 0.05)$, find $P[X \leq 8]$ by

1. direct calculation (Casio calculator is enough)
2. Poission approximation
3. Normal approximation(without half-unit correction)
4. Normal approximation(with half-unit correction)

ex 3.3 answer

1.
$P[X=x]=\binom{100}{x}0.05^x0.95^{100-x}$
The answer is $\sum_{x=0}^{8}\binom{100}{x}0.05^x0.95^{100-x}=0.9369$
2.
$P[X=x]\approx\frac{5^xe^{-5}}{x!}$
The answer is $\sum_{x=0}^{8}\frac{5^xe^{-5}}{x!}=0.9319$
3.
$P[X=8]\approx\Phi(\frac{8-5}{\sqrt{5*0.95}})=\Phi(1.376)=0.916$
4.
$P[X=8]\approx\Phi(\frac{8+0.5-5}{\sqrt{5*0.95}})=\Phi(1.61)=0.9463$

Central Limit Theorem(Lindeberg-Levy)

The most "classic" version of central limit theorem.

$X_1, X_2, \ldots, X_n$ are n i.i.d random variables with $E[X_i] = \mu$, $Var[X_i] = \sigma^2$, then it's easy to see that

1. $E[\sum_{i=1}^{n} X_n] = n\mu$
2. $Var[\sum_{i=1}^{n} X_n] = n\sigma^2$

The central limit theorem claims that as long as n is big enough, the sum $\sum_{i=1}^{n} X_n$ follows a normal distribution with expectation and variance above.

It's important to notice that when n is big enough, the average of $X_1, X_2, \ldots, X_n$ also approximately follows a normal distribution with expectation μ and variance $\frac{\sigma^2}{n}$.

ex 3.4

X is a discrete uniform random variable with range $\{1, 2, \ldots, 6\}$. $X_1, X_2, \ldots, X_{100}$ are 100 i.i.d sample of X. What's the probability that the sum of these 100 numbers falls in the range of $[300, 399]$?

ex 3.4 answer

Use the central limit therorem. Set $M := \sum_{i=1}^{100} X_i$, then $E[M] = 350, Var[M] = 100 * Var[X] = 100 * \frac{35}{12} = 291.67$. M approximately follows $N(350, 17.078^2)$.

$$
\begin{aligned}
&P[300 \leq M \leq 399] \\
=&P[299.5 \leq M \leq 399.5] \\
=&P[-2.957 \leq M \leq 2.898] \\
=&0.9981 - 0.0016 = 0.9965
\end{aligned}
$$

ex 3.4 answer(theoretical value)

By comparation, let's calculate the theoretical value of the probability. The total conditions are 6^{100}, and we need to calculate how many methods can lead to a sum in $[300, 399]$. That is the sum of coefficients from x^{300} to x^{399} in the expression $(x + x^2 + x^3 + x^4 + x^5 + x^6)^{100}$. By mathematica calculation the result is

$$\frac{651143833793694402570344336937171074415138808053670555322180232871923298151381}{653318623500070906096690267158057820537143710472954871543071966369497141477376} = 0.99667$$

Central Limit Theorem(Lyapunov)

This version is the most generalized version. As long as n is big enough, there's even no need for $X_1, X_2, \ldots, X_n$ to be identical distributed. The expectation and variance of each random variable should exist. There're some other requirements, which is not covered in this course.

Why type-A uncertainty is normal distributed

$$u_A = t_{n-1}\frac{S}{\sqrt{n}}$$

Outline

extra topic and Q&A

You're recommended to summarize the properties of exponential distribution, gamma distribution and normal distribution into a table. Also, the sample midterm exam is avaliable on canvas, and now you can finish exercise 2-13.