STA256H5S: Probability and Statistics I

Chapter 4: Continuous Random Variables

Thomas Yoon

Department of Mathematical and Computational Sciences
University of Toronto Mississauga

July 2018

Table of Contents

1. Cumulative Distribution Function
2. Continuous Random Variables
3. Continuous Distributions
    - Uniform Distribution(Continuous case)
    - Normal Distribution
    - Exponential Distribution
    - Gamma distribution
    - Chi-sq Distribution
    - Beta Distribution
4. Markov's Inequality,Chebyshev's Theorem, and Empirical Rule

Table of Contents

1. Cumulative Distribution Function
2. Continuous Random Variables
3. Continuous Distributions
   - Uniform Distribution(Continuous case)
   - Normal Distribution
   - Exponential Distribution
   - Gamma distribution
   - Chi-sq Distribution
   - Beta Distribution
4. Markov's Inequality,Chebyshev's Theorem, and Empirical Rule

UNIVERSITY OF TORONTO MISSISSAUGA

Definition

Let X be a random variable. The cumulative distribution function, CDF (sometimes also called distribution function),$F_X(x) = F(x)$ is defined as follows:

$$F_X(x) = F(x) = P(X \leq x) \text{ for } -\infty \leq x \leq \infty$$

Example

Let X be a random variable with the following probability distribution:

x	0	1	2	3	4
p(x)	0.1	0.2	0.3	0.2	0.2

1. Obtain the cumulative distribution function.
2. Plot $F_X(x)$.

Properties of CDF

Let X be a random variable, and $F_X(x)$ be its CDF. Then following properties are true:

1. $\lim_{x \to -\infty} F_X(x) = 0$
2. $\lim_{x \to \infty} F_X(x) = 1$
3. $F_X(x)$ is a non-decreasing function of x. This means $x_1 < x_2 \implies F_X(x_1) \leq F_X(x_2)$
4. $F_X(x)$ is right continuous. Recall: Right continuous means $\lim_{x \to x_0^+} F(x) = F(x_0) \; \forall x_0 \in \mathbb{R}$

Table of Contents

1. Cumulative Distribution Function
2. Continuous Random Variables
3. Continuous Distributions
   - Uniform Distribution(Continuous case)
   - Normal Distribution
   - Exponential Distribution
   - Gamma distribution
   - Chi-sq Distribution
   - Beta Distribution
4. Markov's Inequality,Chebyshev's Theorem, and Empirical Rule

Definition

Let X be a random variable. If $F_X(x)$ is continuous for $-\infty < x < \infty$ (not just right continuous, but also left continuous), then X is a **continuous** random variable.

Definition

- For continuous random variable X, the function that represents probability distribution is called **probability density function**, PDF, $f_X(x)$ (Recall: the analogue for discrete random variable was probability mass function).
- $f_X(x) = \frac{dF_X(x)}{dx}$ wherever $\frac{dF_X(x)}{dx}$ exists.
- Sometimes PDF is also called density function.

Consideration

Start visualizing probability density as $n+1$ dimensional measure, where $n=$number of random variables in interest. In case of one random variable, the probability density is the second dimensional measure, width. Then, when we integrate the width over all length $\in S$, we get total area of 1.
If we are interested in a function of two random variables (say $Y = X_1 + X_2$), then $X_1 \times X_2$ plane $\in S$ at Z=0 while the probability density is the $2+1$ dimensional measure, height. Then when we integrate the height over the Z=0 plane $\in S$ with respect to the height, we get total volume of 1.
Same logic for functions of higher number of random variables...

Properties of PDF

Let $f_X(x)$ be a probability density function of a continuous random variable X. $f_X(x)$ satisfy following conditions:

- $P(X = x) = 0 \; \forall x \in \mathbb{R}$
- $f_X(x) \geq 0$ for $-\infty < y < \infty$.
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- More over, $\int_{x \in S} f_X(x)dx = 1$ where S is sample space.

Example

Let X be a random variable with following probability density function.

$$f_X(x) = \frac{1}{2} I_{[0,1]}(x) + cI_{(1,4]}(x)$$

using indicator functions, I, or equivalently,

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } 0 \leq x \leq 1 \\ c & \text{for } 1 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

- Find value of c such that $f_X(x)$ is a valid PDF.
- Graph $f_X(x)$ and $F_X(x)$.
- Find $P(0.5 \leq X \leq 2)$ in two ways.

Example- Other way

Consider the following CDF for X.

$$F_X(x) = \frac{x}{8} I_{(0,2)}(x) + \frac{y^2}{16} I_{[2,4)}(x) + I_{[4,\infty)}(x)$$

.

1. Find the probability density function of X, $f_X(x)$.
2. Find $P(X \geq 1.3)$.
3. Find $P(X \leq 4 | X \geq 1.5)$

Definition

The expected value of a continuous random variable X is similar to the expected value of a discrete random variable. The only difference is the use of integral instead of summation.

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

In some cases, $E(X)$ does not exist (i.e., when the integral does not exist), but in this course, no need to worry about this!

Definition

$E(g(X))$ for a continuous random variable X is also very similar to that for discrete random variable.

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Again, provided that the integral exists.
Note: $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{x \in S} x f_X(x) dx$ and similarly,
$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{x \in S} g(x) f_X(x) dx$ as $f_X(x) = 0 \ \forall x \notin S$

Properties of Expected values (We already know most of them.)

Let $c \in \mathbb{R}$ and $g_1(X), g_2(X), ..., g_n(X)$ be real valued functions of continuous random variable X. Then, following statements are true:

1. $E(c) = c$
2. $E(cg_i(X)) = cE(g_i(X))$
3. $E(\sum_{i=1}^{n} g_i(X)) = \sum_{i=1}^{n} E(g_i(X))$. Preserving linearity.

Proof

1.
2.
3.

Example

Let X be a continuous random variable whose $E(X) = \mu, var(X) = \sigma_X^2$ and $a, b \in \mathbb{R}$. Prove the following items:

1. $E(aX + b) = aE(X) + b$
2. $var(aX + b) = a^2\sigma_X^2$

Table of Contents

1. Cumulative Distribution Function
2. Continuous Random Variables
3. Continuous Distributions
   - Uniform Distribution(Continuous case)
   - Normal Distribution
   - Exponential Distribution
   - Gamma distribution
   - Chi-sq Distribution
   - Beta Distribution
4. Markov's Inequality,Chebyshev's Theorem, and Empirical Rule

Definition

X has uniform probability distribution on the interval (θ_1, θ_2) when:

$$X \sim unif(\theta_1, \theta_2) \iff f_X(x) = \frac{1}{\theta_2 - \theta 1} I_{[\theta_1, \theta_2]}(x) \text{ where } \theta_1 < \theta_2$$

note: The term uniform comes from the fact that the width is a constant and it is uniform throughout the region $\theta_1 \leq X \leq \theta_2$

There is also a discrete version of uniform distribution, where the pmf is simply $P(X = a) = \frac{1}{^\# S}$ $\forall a \in S$ where $^\# S$ represents the cardinality of sample space. You can also see that the width is constant and uniform throughout S.

$$X \sim unif(\theta_1, \theta_2) \implies E(X) = \frac{\theta_1 + \theta_2}{2}, \sigma_X^2 = \frac{(\theta_2 - \theta_1)^2}{12}$$

Proof: Will be on Blackboard in a separate pdf document.

Proof

Example

Suppose STA256 test 1 marking distribution followed uniform distribution with minimum score of 25% and maximum score of 93%.

- Sketch the pdf.
- Find the probability that a randomly chosen student receives a B class grade (70-79%)?
- Find the CDF.
- What was the class average of the test?
- Suppose $Y = X^2$. Is Y also a uniform distribution?

Definition

A random variable X follows normal distribution with mean μ and variance σ_X^2 when:

$$X \sim N(\mu, \sigma_X^2) \iff f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(x-\mu)^2}{2\sigma_X^2}} \text{ for } -\infty \leq x \leq \infty$$

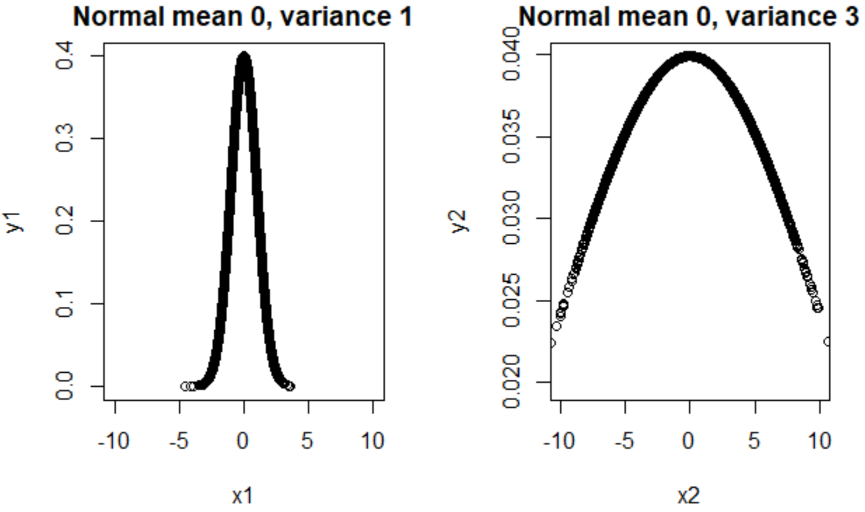
where $-\infty \leq \mu \leq \infty, \sigma_X^2 > 0$

Normal Distribution: Expected value and Variance

$$X \sim N(\mu, \sigma_X^2) \implies E(X) = \mu, var(X) = \sigma_X^2$$

Proof: Will be on Blackboard in a separate pdf.

Plot of normal distributions



Normal Distribution Properties:

When a normal distribution is plotted, you get the famous normal density curve (AKA. bell curve). Notice that the normal distribution can be perfectly explained by just two parameters, μ, σ_X^2 which stands for the mean and the variance of the normal distribution.
Notice that the mean of the normal distribution is at the centre of the normal curve, and the curve is perfectly symmetrical by the vertical axis at the mean. The standard deviation tells how sharp or wide the normal curve will be. The greater the standard deviation, the wider the normal distribution (more probability at the tails), and vice versa.

Special normal distribution: Standard normal distribution

Normal distributions are very common, and as described previously, normal distribution with different parameters have different curves and different probability distribution. So, we came up with a 'standard' normal distribution so we only have to find probabilities of this one normal distribution. It also has a **drawback**: when the original distribution is not normal distribution, you must convert it to standard normal distribution.

Standard normal distribution, $Z \sim N(0,1) \iff f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}$

Standard normal distribution has $E(Z) = 0, var(Z) = \sigma_Z^2 = 1$

Converting from normal to standard normal

Let $X \sim N(\mu, \sigma_X^2)$. Then:

$$Z = \frac{X - \mu}{\sigma_X}$$

where Z is the standard normal distribution.

Example

Canadian men's height follows normal distribution with mean of 178 cm and variance of 169 cm^2. Eldie's height is 151 cm.

1. He would like to know the proportion of Canadian men who are shorter than him. What's the proportion?
2. Eldie's brother is 180 cm. If a random Canadian man is selected, what is the probability that his height is between Eldie's and his brother's?
3. What is the proportion of Canadian men who are taller than Eldie's brother but shorter than 2 standard deviation away from the mean?

Example

Most medical schools in Canada require applicants to write and report their performances on MCAT. MCAT scores are standardized test meaning the results are normally distributed with mean 7 and standard deviation of 2. Lindsey scored 11.5 on physical sciences section.

1. Suppose probability of getting in is completely dependent on applicant's MCAT score. ie., if they scored 50 percentile, the probability of getting in is 50%. What is the probability that Lindsey will get into medical school?
2. How many people scored between 10 and 13?
3. Applicants who scored within bottom 20 percentile are flagged. What is the maximum score at which the student will be flagged?

Normal distribution: Moment Generating Function

Let $X \sim N(\mu_X, \sigma_X^2)$. Then, X has moment generating function,
$M_X(t) = e^{\mu_X t + \frac{\sigma_X^2 t^2}{2}}$

Proof

Example

Let $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Let $X \perp\!\!\!\perp Y$. Show that for $a, b \in \mathbb{R}$, $W = aX = bY$ also follows normal distribution using MGF, and find $E(W), var(W)$.

Example

Wakanda has been separate from the rest of the earth for very long time. Hence, the intelligence of people in Wakanda are very different from the rest. Let Wakandan's IQ points follow normal distribution with mean 130 and variance of 64. In the rest of the world, our IQ points follow normal distribution with mean 100 and variance of 100.

1. Find the proportion of non-Wakandans whose IQ points are higher than 1 standard deviation above the average IQ points of Wakandans.
2. Suppose an offspring has the average IQ between parents. What distribution does the offspring between Wakandan and non-Wakandan follow?
3. What is the probability that the offspring's IQ is between the mean of Wakandan and non-Wakandan?

Relationship between Poisson and Exponential distribution

From Poisson distribution, recall that the probability of at least one event occurring in t number of independent intervals is $1 - (e^{-\lambda})^t = 1 - e^{-\lambda t}$. Consider random variable X be the time to the first Poisson event. This means, the probability that the length of time until the first event will be less than or equal to x is the same as the probability that at least one Poisson event will occur in x. i.e.,
$P(X \leq x) = 1 - e^{-\lambda x} = F_X(x),$ where $x > 0$. Then,
$f_X(x) = \frac{dF_X(x)}{dx} = \lambda e^{-\lambda x}$.

UNIVERSITY OF TORONTO MISSISSAUGA

Definition: Exponential distribution

There are two ways of writing exponential distribution. You may use rate parameter, $\lambda > 0$, or scale parameter, $\beta > 0$.

$$X \sim Exp(\lambda) \iff f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)(x)}$$

Alternatively, setting $\lambda = \frac{1}{\beta}$,

$$X \sim Exp(\beta) \iff f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} I_{[0,\infty)(x)}$$

In this class, we will use the scale parameter notation more often.

Exponential distribution: Expected value and variance

$X \sim Exp(\beta) \implies E(X) = \beta, var(X) = \beta^2$ Proof: Will be posted on Blackboard in a separate pdf document.

Exponential distribution: MGF of Exponential distribution

$$X \sim Exp(\beta) \implies M_X(t) = (1 - \beta t)^{-1}$$

Proof

Example

Let $X \sim Exp$ and $P(X > 2) = 0.0821$.

1. Find β
2. Find $P(1 \leq 1.7)$

Example

Michelle is counting the number of students entering a bar on St. Patrick's day. The students' arrivals follow exponential distribution with mean 100 per hour.

1. What is the probability that more than 150 students arrive in an hour?
2. If there are too many students entering at the same time ($> c$), then they fight and the management has to call the cops. What is the maximum number of students before they need t o call the cops if cops are called 1% of the time?

UNIVERSITY OF TORONTO MISSISSAUGA

Definition: Gamma distribution

A random variable X follows gamma distribution with parameters, $\alpha > 0, \beta > 0$ when:

$$X \sim \Gamma(\alpha, \beta) \iff f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}(e)^{\frac{-x}{\beta}} I_{[0,\infty)}(x)$$

where Gamma function, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$

Properties of Gamma function

- Visualize the gamma function as the normalizing constant of gamma distribution,$\Gamma(\alpha, \beta = 1)$. This means that $\forall X \sim \Gamma(\alpha, 1),\ 1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\Gamma(\alpha)} x^{\alpha-1}(e)^{-x} I_{[0,\infty)}(x)dx = \frac{\int_{-\infty}^{\infty} x^{\alpha-1}(e)^{-x} I_{[0,\infty)}(x)dx}{\Gamma(\alpha)} = \frac{\int_{-\infty}^{\infty} x^{\alpha-1}(e)^{-x} I_{[0,\infty)}(x)dx}{\int_0^{\infty} x^{\alpha-1}(e)^{-x}dx}$
- Normalizing constant means the constant makes the integral perfect pdf, i.e., the integral from $-\infty$ to ∞ becomes 1 after multiplying by the normalizing constant.

Properties of Gamma function

- $\forall n \in \mathbb{N}, \Gamma(n) = (n-1)!$
- There are lots of particular gamma function values, but pay special attention to $\Gamma(\frac{1}{2}) = \sqrt{\pi}, \Gamma(\frac{3}{2}) = \frac{\pi}{2}, \Gamma(1) = 0! = 1$
- $\forall n \in \mathbb{R}, \Gamma(n) = (n-1)\Gamma(n-1)$

Proof of 3rd point

Use: Integral by parts:

$$\int udv = uv - \int vdu$$

where $u = x^{\alpha-1}, dv = e^{-x}dx$

Gamma distribution: Expected values and Variance

$$X \sim \Gamma(\alpha, \beta) \implies \mu = E(X) = \alpha\beta, \sigma_X^2 = \alpha\beta^2$$

Proof: Will be provided on Blackboard in a separate pdf document.

Definition

Let $n \in \mathbb{N}$. Random variable, X has chi-square distribution, with n degrees of freedom $\chi_n \iff$ X is Gamma distributed with $\alpha = \frac{n}{2}, \beta = 2$

Example

Suppose that a random variable X has a probability density function $f_X(x) = cx^3 e^{\frac{-x}{2}} I_{(0,\infty)}(x)$.

1. Find the normalizing constant k such that $f_X(x)$ is a valid pdf.
2. Does X have chi-sq distribution? If so, with what degrees of freedom?
3. What are the mean and the variance of X? (hint: Use the fact that it is also a gamma distribution)

Definition

$$X \sim \chi_n \implies E(X) = n, \sigma_X^2 = var(X) = 2n$$

Proof

Using the fact that

$$\chi_n \iff \Gamma(\frac{n}{2}, 2)$$

Definition

X follows beta distribution with parameters, $\alpha > 0, \beta > 0$:

$$X \sim Beta(\alpha, \beta) \iff f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{[0,1]}(x)$$

Trivially, $X \sim Beta(1, 1) \iff X \sim unif(0, 1)$. Can you see it directly?

Beta distribution: Expected value and Variance

$$X \sim Beta(\alpha, \beta) \implies \mu = \frac{\alpha}{\alpha + \beta}, var(X) = \sigma_X^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Proof: will be provided on Blackboard on a separate pdf document.

Similar Example

Let X be a continuous random variable with pdf:

$$f_X(x) = cx^3(1-x)^2$$

1. Find the value of normalizing constant, c such that $f_X(x)$ is a valid pdf.
2. Find $E(X), \sigma_X^(2)$

Example

Proportion of attendance in Drake's concerts follow beta distribution with $\alpha = 2, \beta = 2$ i.e., tickets are sold out when x=1, not a single ticket is sold when x=0, half of tickets are sold when x=0.5, etc...

1. What is the probability that in a randomly selected concert, the attendance is less than 0.75?
2. What is the mean and variance of the proportion of attendance?

Definition (MGF Review)

Recall from previous lecture that MGF in discrete case was
$M_X(t) = E(e^{tX}) = \sum_{\text{all } x} e^{tx} P(X = x)$ and
$M_{g(X)}(t) = E(e^{tg(X)}) = \sum_{\text{all } x} e^{tg(x)} P(X = x)$.
In continuous case, you change sums to integrals.
$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$
$M_{g(X)}(t) = E(e^{tg(X)}) = \int_{-\infty}^{\infty} e^{tg(x)} f_X(x) dx$

Example

We know: $X \sim N(\mu_X, \sigma_X^2) \implies Z = \frac{X-\mu_X}{\sigma_X} \sim N(0,1)$.
Prove it using MGF (Follow below).

1. Find MGF for Z.
2. Recall first moment and second moment yields $E(X), E(X^2)$ respectively. Find $E(Z), \sigma_Z^2$

Table of Contents

1. Cumulative Distribution Function
2. Continuous Random Variables
3. Continuous Distributions
   - Uniform Distribution(Continuous case)
   - Normal Distribution
   - Exponential Distribution
   - Gamma distribution
   - Chi-sq Distribution
   - Beta Distribution
4. Markov's Inequality,Chebyshev's Theorem, and Empirical Rule

Definition: Markov's Inequality

If X is a random variable (can be continuous or discrete) that takes only non-negative values (i.e., $x \geq 0$). Then, $\forall a > 0, E(X) \geq aP(X \geq a)$.

Proof: Continuous Case (Discrete case proof for exercise)

Let X be a random variable with $|\mu_X| < \infty$ and $\sigma_X^2 > 0$. Then:

$$\forall k > 0, P(|X - \mu_X| < k\sigma_X) \geq 1 - \frac{1}{k^2}$$

or equivalently:

$$\forall k > 0, P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

Proof

Let $Y = (X - \mu_X)^2$. Then, notice that Y takes only non-negative values, and we can use Markov's inequality with $a = k^2\sigma_X^2$. Markov's Inequality tells us:

$$k^2\sigma_X^2 P((X - \mu_X)^2 \geq k^2\sigma_X^2) \leq E((X - \mu_X)^2)$$

Notice $(X - \mu_X)^2 \geq k^2\sigma_X^2$ only occurs when $|X - \mu_X| \geq k\sigma_X$, above equation becomes:

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{E((X - \mu_X)^2)}{k^2\sigma_X^2} = \frac{1}{k^2}$$

as we know $E((X - \mu_X)^2)$ is σ_X^2 by definition of variance.

Definition: Empirical Rule

Empirical rule states that if $X \sim N(\mu, \sigma_X^2)$, then approximately:
68% of the observations are between $1\sigma_X$ away from the mean.
95% of the observations are between $2\sigma_X$ away from the mean.
99.7% of the observations are between $1\sigma_X$ away from the mean.
Since Empirical Rule only applies to normal distributions, Chebyshev's theorem is a lot stronger statement. i.e., *Chebyshev* $\not\Longrightarrow$ *Empirical*, but *Empirical* $\Longrightarrow$ *Chebyshev*

Example

Let $X \sim exp(\beta)$.

1. Find $P(|X - \mu| \leq 2\sigma_X)$
2. Find the exact same quantity according to Empirical Rule.
3. Find the exact same quantity using Chebyshev's theorem.