13 Continuous Distributions

Computer Science Mentors 70

Nov 30 - Dec 4
Fall 2020

Prepared by: Aishani Sil, Austin Lei, Agnibho Roy, Debayan Bandyopadhyay, Abinav Routhu

Roast us here: *https://tinyurl.com/csm70-feedback20*

Discrete vs Continuous Probability

Here is a table illustrating the parallels between discrete and continuous probability.

Discrete	Continuous
$P[X=k] = \sum_{\omega \in \Omega : X(\omega)=k} P(\omega)$	$P[k < X \leq k + \mathrm{d}x] = f_X(k)\mathrm{d}x$ (*)
$P[X \leq k] = \sum_{\omega \in \Omega : X(\omega) \leq k} P(\omega)$	$P[X \leq k] = F_X(k)$
$E[X] = \sum_{a \in A} a \cdot P[X=a]$	$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$
$E[\phi(X)] = \sum_{a \in A} \phi(a) \cdot P[X=a]$	$E[\phi(X)] = \int_{-\infty}^{\infty} \phi(x) \cdot f_X(x) dx$
$\sum_{\omega \in \Omega} P[\omega] = 1$	$\int_{-\infty}^{\infty} f_X(x) dx = 1$

(*) When solving problems with continuous distributions, you can think of $f_X(k)$ as being analogous to $P[X = k]$ in discrete distributions, but they are not equal.

I Intro to Continuous Distributions

1. **PDFs**

Consider the following functions and determine whether or not they are valid probability density functions.

(a) $f(x) = \sin(x)$

Solution: This is not valid because $\sin(x)$ can be negative.

(b) $f(x) = x$ for $0 \leq x \leq 1$, and $f(x) = 0$ everywhere else.

Solution: This is not valid, since

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^1 x\mathrm{d}x = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2} \neq 1$$

(c) $f(x) = 1$ for $0 \leq x \leq 1$, and $f(x) = 0$ everywhere else.

> **Solution:** This is valid, since $f(x) \geq 0$ for all x, and
>
> $$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^1 1\mathrm{d}x = [x]_0^1 = 1 = 1.$$

(d) $f(x) = e^{-x}$ for $x \geq 0$, and $f(x) = 0$ everywhere else.

> **Solution:** This is valid, since $f(x) \geq 0$ for all x, and
>
> $$\int_{-\infty}^{\infty} e^{-x}\mathrm{d}x = \int_0^{\infty} e^{-x}\mathrm{d}x = [-e^{-x}]_0^{\infty} = 0 - (-1) = 1.$$
>
> (This is the pdf of a Poisson(1) distribution.)

2. **Disk**
Define a continuous random variable R as follows: we pick a point uniformly at random on a disk of radius 1; the value of R is distance of this point from the center of the disk. We will find the probability density function of this random variable.

(a) Why is R not $U(0, 1)$?

> **Solution:** We can think of it somewhat in areas. There are more points that have larger radius than smaller radius, and since the likelihood of selecting any particular point is equal, it is more likely to get a larger radius than a smaller one.

(b) What is the probability that R is less than r, for any $0 \leq r \leq 1$? What is the CDF $F_R(r)$ of the random variable R?

> **Solution:** r^2, because the area of the circle with distance between 0 and r is πr^2, and the area of the entire circle is π.
> Thus, we have that $F_R(r) = r^2$ for $0 \leq r \leq 1$.

(c) What is the PDF $f_R(r)$ of the random variable R?

> **Solution:** By definition,
>
> $$f_R(r) = \frac{\mathrm{d}}{\mathrm{d}r} F_R(r) = \frac{\mathrm{d}}{\mathrm{d}r} r^2 = 2r$$
>
> for $0 \leq r \leq 1$.

(d) Now say that $R \sim U(0, 1)$. Are you more or less likely to hit closer to the center than before?

> **Solution:** More likely. Let's evaluate the probability that $R \leq c, c \in (0, 1)$ in both cases. In the first case, $P(R \leq c) = c^2$. In the second case, $P(R \leq c) = c$. For $c \in (0, 1), c \geq c^2$.

3. **Joint Density**
The joint density for the random variables X and Y is defined by $f(x, y) = x + 2y$ for $0 \leq x, y \leq c$ for some positive real number c, and $f(x, y) = 0$ for all other (x, y).

(a) For what value of c is this a valid joint density?

Solution: For the joint density to be valid, we need that the total integral equals 1, so

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_0^c\int_0^c x + 2y\mathrm{d}x\mathrm{d}y \\
&= \int_0^c \left[\frac{x^2}{2} + 2xy\right]_0^c \mathrm{d}y \\
&= \int_0^c \frac{c^2}{2} + 2cy\mathrm{d}y \\
&= \left[\frac{c^2}{2}y + cy^2\right]_0^c \\
&= \frac{3c^3}{2},
\end{aligned}$$

so we have that $c = \sqrt[3]{\frac{2}{3}}$.

(b) Compute $P(X < Y)$.

Solution: To get $P(X < Y)$, we need to integrate the pdf over the region where $X < Y$, so

$$\begin{aligned}
P(X < Y) &= \int_0^c\int_0^y f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_0^{\sqrt[3]{\frac{2}{3}}}\int_0^y x + 2y\mathrm{d}x\mathrm{d}y \\
&= \int_0^{\sqrt[3]{\frac{2}{3}}} \left[\frac{x^2}{2} + 2xy\right]_0^y \mathrm{d}y \\
&= \int_0^{\sqrt[3]{\frac{2}{3}}} \frac{5y^2}{2}\mathrm{d}y \\
&= \left[\frac{5y^3}{6}\right]_0^{\sqrt[3]{\frac{2}{3}}} \\
&= \frac{5}{9}.
\end{aligned}$$

(c) Compute $E[X|Y = y]$ for $0 \leq y \leq c$.

Solution: We have that

$$\begin{aligned} E[X|Y=y] &= \int_0^c f(x,y)\mathrm{d}x \\ &= \int_0^{\sqrt[3]{\frac{2}{3}}} x + 2y\mathrm{d}x \\ &= \left[\frac{x^2}{2} + 2xy\right]_{x=0}^{\sqrt[3]{\frac{2}{3}}} \\ &= 2\sqrt[3]{\frac{2}{3}}y + \frac{\sqrt[3]{\frac{4}{9}}}{2} \end{aligned}$$

(d) Compute $E[XY]$.

Solution: We have that

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_0^c\int_0^c x^2y + 2xy^2\mathrm{d}x\mathrm{d}y \\ &= \int_0^c \left[\frac{x^3y}{3} + x^2y^2\right]_0^c \mathrm{d}y \\ &= \int_0^c \frac{c^3y}{3} + c^2y^2\mathrm{d}y \\ &= \left[\frac{c^3y^2}{6} + \frac{c^2y^3}{3}\right]_0^c \\ &= \frac{c^5}{2} \\ &= \frac{\left(\sqrt[3]{\frac{2}{3}}\right)^5}{2} \\ &= \frac{\sqrt[3]{\frac{4}{9}}}{3}. \end{aligned}$$

4. **Transformation of Densities**
Let X be a Exponential(λ) random variable, where $\lambda > 0$, and define random variable $Y = X^{\lambda}$. We wish to compute the pdf of Y.

(a) What is the cdf $F_X(x)$ of X?

Solution: By definition, for $x \geq 0$,

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_0^x \lambda e^{-\lambda t} \mathrm{d}t \\ &= \left[-e^{-\lambda t}\right]_0^x \\ &= 1 - e^{-\lambda x}, \end{aligned}$$

and the cdf will be 0 for $x < 0$.

(b) What is the cdf $F_Y(y)$ of Y? (Hint: How can you relate F_Y to F_X?)

Solution: By definition, for $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^\lambda \leq y) \\ &= P(X \leq y^{\frac{1}{\lambda}}) \\ &= F_X(y^{\frac{1}{\lambda}}) \\ &= 1 - e^{-\lambda y^{\frac{1}{\lambda}}} \end{aligned}$$

and the cdf will be 0 for $y < 0$.

(c) What is the pdf $f_Y(y)$ of Y?

Solution: We have that

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) \\ &= \frac{\mathrm{d}}{\mathrm{d}y} 1 - e^{-\lambda y^{\frac{1}{\lambda}}} \\ &= \frac{1}{\lambda} \lambda y^{\frac{1}{\lambda}-1} e^{-\lambda y^{\frac{1}{\lambda}}} \\ &= y^{\frac{1}{\lambda}-1} e^{-\lambda y^{\frac{1}{\lambda}}} \end{aligned}$$

II Uniform, Exponential, and Normal Distributions

5. **Exponential/Normal Intro**

(a) There are certain organisms that don't age called hydra. The chances of them dying is purely due to environmental factors, which we will call λ. On average, 2 hydra die within 1 day. What is the probability you have to wait at least 5 days before a hydra dies?

Solution: $\lambda = 2, X \sim Exp(2)$
$P(X \geq 5) = \int_5^\infty \lambda e^{-\lambda x} dx = \int_5^\infty 2e^{-2x} dx = -e^{-2x}|_5^\infty = e^{-10} = \frac{1}{e^{10}}$

(b) There are on average 8 office hours in a day. The scores of an exam followed a normal distribution with an average of 50 and standard deviation of 6. If a student waits until an office hour starts, what is the expected value of the sum of the time they wait in hours and their score on the exam?

Solution: Model the beginnings of office hours as 8 points throughout a 24-hour day. Model the student's arrival as 1 point in a 24-hour day. Then in total there are 9 points distributed throughout a 24-hour day. By symmetry, the expected time interval from the student's arrival to the next office hour is $\frac{24}{9}$.

$$\text{E(waiting time)} = \frac{24}{9}$$
$$\text{E(score)} = \text{E}(\mathcal{N}(50, 36)) = 50$$

By linearity of expectation, the sum is $52\frac{2}{3}$.

6. **Recruiting Season**
You have a phone interview with a company, and you read a strange review on Glassdoor indicating that the length of these interviews follow an exponential distribution with a mean of 20 minutes.

(a) What is the variance of X, the time an interview lasts for?

Solution: Let us define a random variable X to be the amount of time your interview lasts. Since the mean is 20 minutes, we know that $X \sim \text{Exp}(\frac{1}{20})$. Let us compute the variance by finding $E[X^2] - E[X]^2$.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^{\infty} x^2 \frac{1}{20} e^{-\frac{x}{20}} dx = \left[-x^2 e^{-\frac{x}{20}}\right]_0^{\infty} + \int_0^{\infty} 2xe^{-\frac{x}{20}} dx = 0 + \frac{2}{\frac{1}{20}} E[X] = 800.$$

$$E[X]^2 = \frac{1}{(\frac{1}{20})^2} = 400 \longrightarrow \text{Var}(X) = 800 - 400 = 400 \text{ minutes.}$$

You've likely gone over the formula for variance of an exponential distribution as $\frac{1}{\lambda^2}$, so plugging into there works too :)

(b) What is the probability that your interview will last at most 10 minutes?

Solution: This is the CDF evaluated at 10. From integrating the PDF, we arrive at $F_X(x) = 1 - e^{-\frac{x}{20}} \longrightarrow F_X(10) = 0.39$.

(c) You are now in the middle of the interview and it has been going on for 600 minutes! What is the probability that the interview last longer than 10 *more* minutes?

Solution: Remember that the exponential distribution is memoryless, meaning that the length of time in which the interview has not yet ended has no bearing on how much longer it will take. More concretely,

$$P[X > 610 \mid X > 600] = P[X > 10].$$

Thus, this is just $1 - P[X \leq 10]$. We found this probability in part (b), so our final answer is $1 - 0.39 = 0.61$.

7. **Penguins!**
Professor Sahai decides that he wants to vacation but wants to do so in isolation due to the coronavirus so he ventures to Antarctica. He read once that penguins have a height anywhere from 3ft to 5ft with uniform probability, but is skeptical so

decides to see for himself. We want to see how closely the average of the penguins he measures is to the true average. Let:

$$\hat{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

be the average of the n height samples from the population of penguins, each independently and randomly collected.

(a) Calculate the expected value and variance for the height of a penguin

Solution: We calculate the expected value of the sample and take the distribution of the height to be uniform. Thus, $E[X_i] = 4$ and $Var[X_i] = E[X^2] - E[X]^2 = \int_3^5 x^2 \frac{1}{2} dx = \frac{49}{3} - 4^2 = \frac{1}{3}$

$$E[\hat{X}] = E[\frac{X_1 + X_2 + \cdots + X_n}{n}] = \frac{1}{n} E[X_1 + X_2 + \cdots + X_n] = \frac{1}{n} n E[X_1] = 4$$
$$Var[\hat{X}] = Var[\frac{X_1 + X_2 + \cdots + X_n}{n}] = \frac{1}{n^2} Var[X_1 + X_2 + \cdots + X_n] = \frac{1}{n^2} n Var[X_1] = \frac{1}{3n}$$

(b) Calculate a 95% confidence interval for the average height of the penguins for an arbitrary n using Chebyshev's Inequality. Interpret this interval.

Solution: Using Chebyshev's inequality, we find that

$$Pr(|\hat{X} - \mu| \geq c) \leq \frac{Var(X)}{c^2} = \frac{1}{3nc^2} = 0.05$$

Solution. Solving for c, we get $c = \sqrt{\frac{20}{3n}}$. The interpretation of this is that the probability that our sample deviates from more than this value of c from the true mean is only 5% of the time, meaning that the following is a 95% confidence interval:

$$\hat{X} \in (4 - \sqrt{\frac{20}{3n}}, 4 + \sqrt{\frac{20}{3n}})$$

(c) Calculate a 95% confidence interval for the average height of the penguins for an arbitrary n using CLT. You may assume that n is sufficiently large. You may assume that $Pr(-1.96 < \mathcal{N}(0,1) < 1.96) = 0.95$

Solution: Since n is sufficiently large, we can approximate $\bar{X} \sim \mathcal{N}(4, \frac{1}{3n})$. Normalizing this would give the standard normal Z function.

$$Pr(-1.96 < \sqrt{3n}(\hat{X} - 4) < 1.96) = 0.95$$
$$Pr(\frac{-1.96}{\sqrt{3n}} + 4 < \hat{X} < \frac{1.96}{\sqrt{3n}} + 4) = 0.95$$

Ths implies that the 95% confidence interval is:

$$\hat{X} \in (4 - \frac{1.96}{\sqrt{3n}}, 4 + \frac{1.96}{\sqrt{3n}})$$

(d) Which of the methods provides a tighter interval for any value of n? What additional conditions do you need to be able to use CLT?

Solution: $1.96 < \sqrt{20}$ so the CLT provides a better bound. In order to perform CLT, we have to draw a large sample randomly from a population, and each sample must be independent.

8. **Which Bound is Strongest?**
Recall the following situation from Week 11: Leanne has a weighted coin that shows up heads with probability $\frac{4}{5}$ and tails with probability $\frac{1}{5}$. Leanne flips the coin 100 times, and the random variable X represents the average number of coins that show up heads. We showed previously that $E[X] = \frac{4}{5}$ and $\text{Var}(X) = \frac{1}{625}$. We now compare the strength of the bounds given by Markov, Chebyshev, and the CLT.

(a) Using Markov's Inequality, determine an upper bound for the probability that $X > 0.9$.

Solution: Using Markov's Inequality is valid since $X \geq 0$. By Markov's Inequality, we have that

$$P(X > 0.9) \leq \frac{E[X]}{0.9} = \frac{8}{9}.$$

(b) Using Chebyshev's Inequality, determine an upper bound for the probability that $X > 0.9$.

Solution: By the Chebyshev's Inequality, we have that

$$\begin{aligned}
P(X > 0.9) &= P(X - 0.8 > 0.1) \\
&\leq P(|X - 0.8| > 0.1) \\
&= P(|X - E[X]| > 0.1) \\
&\leq \frac{\text{Var}(X)}{0.1^2} \\
&= \frac{\frac{1}{625}}{\frac{1}{100}} \\
&= \frac{4}{25}.
\end{aligned}$$

(c) Using the CLT, determine an upper bound for the probability that $X > 0.9$. You may use the fact that $\Phi(2.5) \approx .994$, where Φ is the CDF for the standard normal distribution.

Solution: By the CLT, we may approximate the distribution of X as a Normal distribution with mean $\frac{4}{5}$ and variance $\frac{1}{625}$, or equivalently standard deviation $\frac{1}{25}$. Thus, we have that

$$\begin{aligned}
P(X > 0.9) &= P(X - 0.8 > 0.1) \\
&= P\left(\frac{X - 0.8}{\frac{1}{25}} > \frac{0.1}{\frac{1}{25}}\right) \\
&= P\left(\frac{X - 0.8}{\frac{1}{25}} > 2.5\right) \\
&\approx 1 - \Phi(2.5) \\
&\approx 0.006.
\end{aligned}$$

(d) How do the bounds compare?

Solution: For this problem, Markov < Chebyshev < CLT. Generally, this will hold true.
The Markov Inequality requires a nonnegative random variable and provides the weakest bounds when considering values far from the mean, but can provide better bounds when considering values close to the mean. The Chebyshev Inequality is generally stronger than the Markov Inequality, but is a 2-tailed inequality rather than a 1-tailed inequality, and requires computation of the variance. The CLT usually provides the strongest bounds, but can only be applied in specific situations.

9. **To the Max!**
Let us say that we are drawing random numbers from a uniform continuous distribution $U[0, b]$, but we are not quite sure what the upper bound of this distribution is, denoted by the variable b. We choose to draw samples from this distribution, and take the maximum of these samples. Let X_i be a single sample drawn from the distribution $U[0, b]$. We can represent

$$M = \frac{n+1}{n} \max\{X_1 \ldots X_n\}$$

Will M give us the correct value for b, in expectation? (*hint*: Take the expectation of M. What do we want this expectation to be?)

Solution: Using linearity of expectation, we have:

$$E[M] = E\left[\frac{n+1}{n}\left(\max\{X_1 \ldots X_n\}\right)\right] = \left(\frac{n+1}{n}\right) E[\max\{X_1 \ldots X_n\}]$$

We start with the CDF of the maximum and then derive it to get the PDF, and then calculate the expectation using integration:

$$\begin{aligned}
f_{\max\{X_i\}}(z) &= \frac{d}{dz} F_{\max\{X_i\}}(z) \\
&= \frac{d}{dz} P(X_1 < z)P(X_2 < z)\ldots P(X_n < z) \\
&= \frac{d}{dz} (\frac{z}{b})^n \\
&= \frac{n}{b} (\frac{z}{b})^{n-1}
\end{aligned}$$

We know find the expectation of the max distribution:

$$\begin{aligned}
E[\max\{X_1, X_2, \ldots, X_n\}] &= \int_0^b z \frac{n}{b} (\frac{z}{b})^{n-1} dz \\
&= \frac{n}{b^n} \int_0^b z^n dz \\
&= \frac{n}{b^n} \frac{z^{n+1}}{n+1}\Big|_0^b \\
&= \frac{n}{b^n} \frac{b^{n+1}}{n+1} \\
&= \frac{n}{n+1} b
\end{aligned}$$

Thus, we have

$$E[M] = \left(\frac{n+1}{n}\right)\left(\frac{n}{n+1}\right) b = b$$

This gives us the value that we are trying to estimate, b, in expectation