12 Continuous Random Variables and Distribution, CLT, LLSE

Computer Science Mentors 70

Apr 24 - Apr 28
Spring 2023

Amogh Gupta, Sylvia Jin, Aekus Bhathal, Abinav Routhu, Debayan Bandyopadhyay, Grant Zhao

1 Continuous RV and Distribution

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.

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 $f(x,y) = \frac{2}{3}x + \frac{4}{3}y$ 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

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

so we have that $c = 1$.

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

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

$$\begin{aligned}P(X<Y) &= \int_0^c\int_0^y f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \frac{2}{3}\int_0^1\int_0^y x + 2y\mathrm{d}x\mathrm{d}y \\ &= \frac{2}{3}\int_0^1 \left[\frac{x^2}{2} + 2xy\right]_0^y \mathrm{d}y \\ &= \frac{2}{3}\int_0^1 \frac{5y^2}{2}\mathrm{d}y \\ &= \frac{2}{3}\left[\frac{5y^3}{6}\right]_0^1 \\ &= \frac{5}{9}.\end{aligned}$$

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

Solution: We first need to compute $f_Y(y)$. We have that

$$\begin{aligned}f_Y(y) &= \int_0^c f(x,y)\mathrm{d}x \\ &= \frac{2}{3}\int_0^1 x + 2y\mathrm{d}x \\ &= \frac{2}{3}\left[\frac{x^2}{2} + 2xy\right]_0^1 \\ &= \frac{4}{3}y + \frac{1}{3}.\end{aligned}$$

Now, we have that

$$\begin{aligned}
E[X|Y=y] &= \int_0^c xf(x|Y=y)\mathrm{d}x \\
&= \int_0^1 \frac{xf(x,y)}{f_Y(y)}\mathrm{d}x \\
&= \frac{2}{3}\int_0^1 \frac{x(x+2y)}{\frac{4}{3}y+\frac{1}{3}}\mathrm{d}x \\
&= \frac{2}{4y+1}\int_0^1 x^2+2xy\mathrm{d}x \\
&= \frac{2}{4y+1}\left[\frac{x^3}{3}+x^2y\right]_0^1 \\
&= \frac{6y+2}{12y+3}.
\end{aligned}$$

(d) Compute $E[XY]$.

Solution: We have that

$$\begin{aligned}
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\mathrm{d}x\mathrm{d}y \\
&= \frac{2}{3}\int_0^c\int_0^c x^2y+2xy^2\mathrm{d}x\mathrm{d}y \\
&= \frac{2}{3}\int_0^1 \left[\frac{x^3y}{3}+x^2y^2\right]_0^1 \mathrm{d}y \\
&= \frac{2}{3}\int_0^1 \frac{y}{3}+y^2\mathrm{d}y \\
&= \frac{2}{3}\left[\frac{y^2}{6}+\frac{y^3}{3}\right]_0^1 \\
&= \frac{2}{3}\cdot\frac{1}{2} \\
&= \frac{1}{2}.
\end{aligned}$$

2 Normal and Exponential Distributions, CLT

1. **Exponential Intro**

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

2. **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 $1 - 0.39 = 0.61$.

3. **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 $Var[X_i] = E[X^2] - E[X]^2 = \int_3^5 x^2 \frac{1}{2} dx = \frac{49}{3} - 4^2 = \frac{1}{3}$
>
> $$E[\hat{X}] = E[\frac{X_1 + X_2 + \cdots + X_n}{n}] = \frac{1}{n} E[X_1 + X_2 + \cdots + X_n] = \frac{1}{n} nE[X_1] = 4$$
> $$Var[\hat{X}] = Var[\frac{X_1 + X_2 + \cdots + X_n}{n}] = \frac{1}{n^2} Var[X_1 + X_2 + \cdots + X_n] = \frac{1}{n^2} nVar[X_1] = \frac{1}{3n}$$

(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.

4. **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 approximation 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.

3 Linear Least-Squares Estimation

4 LLSE

1. **Linear Least Squares Estimate: Derivation**
   The LLSE of Y given X, denoted $L[Y|X]$, is the linear estimator $\hat{Y} = g(X) = a + bX$ that minimizes least-squares error:

$$C(g) = E(|Y - g(X)|^2) = E(|Y - a - bX|^2).$$

It turns out $L[Y|X] = E(Y) + \frac{cov(X,Y)}{var(X)}(X - E(X))$.
Let's try to derive this.

(a) Write $C(g)$ as linear function of $E(Y^2)$, $E(X^2)$, $E(Y)$, $E(X)$ and $E(XY)$

Solution:

$$\begin{aligned} C(g) &= E(|Y-a-bX|^2) = E((Y-a-bX)(Y-a-bX)) \\ &= E(Y^2 + a^2 + b^2X^2 - 2aY - 2bYX + 2abX) \\ &= E(Y^2) + a^2 + b^2E(X^2) - 2aE(Y) - 2bE(YX) + 2abE(X). \end{aligned}$$

(b) Using calculus, find the values of a and b that minimize the expression in part a. To simplify the calculation use

$$Cov(X,Y) = E(YX) - E(Y)E(X) \text{ and } Var(X) = E(X^2) - E(X)^2.$$

Solution: To find the values of a and b that minimize that expression, we set to zero the partial derivatives with respect to a and b. This gives the following two equations:

$$\begin{aligned} 0 &= \frac{d}{da}C(g) = 2a - 2E(Y) + 2bE(X) \\ 0 &= \frac{d}{da}C(g) = 2bE(X^2) - 2E(YX) + 2aE(X) \end{aligned}$$

From equation 1 we can find a

$$a = E(Y) - bE(X)$$

Substituting that in equation 2 gives

$$\begin{aligned} 0 &= 2bE(X^2) - 2E(YX) + 2aE(X) \\ &= 2bE(X^2) - 2E(YX) + 2[E(Y) - bE(X)]E(X) \\ &= bE(X^2) - E(YX) + E(Y)E(X) - bE(X)^2 \\ &= b[E(X^2) - E(X)^2] - [E(YX) - E(Y)E(X)] \\ &= bVar(X) - Cov(X,Y) \\ b &= \frac{Cov(X,Y)}{Var(X)} \end{aligned}$$

So finally we get values a and b as

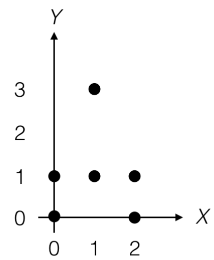
$$\begin{aligned} a &= E(Y) - \frac{Cov(X,Y)}{Var(X)}E(X) \\ b &= \frac{Cov(X,Y)}{Var(X)} \end{aligned}$$

(c) Construct $L[Y|X]$ using the values you found for a and b.

Solution:

$$
\begin{aligned}
L[Y|X] &= a + bX \\
&= E(Y) - \frac{Cov(X,Y)}{Var(X)}E(X) + \frac{Cov(X,Y)}{Var(X)}X \\
&= E(Y) + \frac{cov(X,Y)}{var(X)}(X - E(X))
\end{aligned}
$$

2. The figure below shows the six equally likely values of the random pair (X, Y).



Specify the functions of:

- $L[Y \mid X]$
- $E(X \mid Y)$
- $L[X \mid Y]$
- $E(Y \mid X)$

Solution: Let's calculate some useful properties of the distribution first and then see how we can use them to calculate the estimates.

$$
\begin{aligned}
|\Omega| = 6 &\implies P[\text{one point}] = \frac{1}{6} \\
\mathrm{E}(X) &= 0\left(\frac{2}{6}\right) + 1\left(\frac{2}{6}\right) + 2\left(\frac{2}{6}\right) \\
&= 1 \\
\mathrm{E}(Y) &= 0\left(\frac{2}{6}\right) + 1\left(\frac{3}{6}\right) + 30\left(\frac{1}{6}\right) \\
&= 1 \\
\mathrm{E}(XY) &= 0\left(\frac{3}{6}\right) + 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) \\
&= 1 \\
\mathrm{Cov}(X,Y) = 0 &\implies \mathrm{L}[Y|X] = E(Y)
\end{aligned}
$$

- $L[Y \mid X]$: Using the LLSE formula: $\mathrm{L}[Y \mid X] = \mathrm{E}[Y] + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathrm{E}[Y]) = \mathrm{E}[Y]$. Therefore $\boxed{L[Y \mid X] = 1}$
- $E[X \mid Y]$: Notice the symmetry across $X = 1$. For all values of y, $E[X|Y = y]$ is the same; therefore $\boxed{E[X|Y] = E[X] = 1}$.
- $L[X \mid Y]$: The MMSE estimator for X given Y is a linear function, therefore $\boxed{\mathrm{L}[X \mid Y] = \mathrm{E}[X \mid Y] = 1}$

- $E[Y \mid X]$ For this one we can't make use of symmetry or directly apply what we calculated above. We must go back to the definition of conditional expectation. We can calculate $\mathrm{E}[Y \mid X = x]$ for every point x, and that entirely defines the expression:

$$\mathrm{E}(Y \mid X = x) = \begin{cases} \frac{1}{2} & \text{if } x = 0 \\ 2 & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 2 \end{cases}$$

The above equation is sufficient, but we can go further by realizing that these points are part of a flipped absolute value function centered around $x = 1$: $\boxed{\mathrm{E}[Y \mid X] = \frac{-3}{2}|X - 1| + 2}$. Indeed, this is not linear, which is why $\mathrm{L}[Y \mid X] \neq \mathrm{E}[Y \mid X]$.