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

$f_X(x)\,dx = Pr[X=x]$

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

NO

$\sin\left(\frac{3\pi}{2}\right) = -1 < 0$

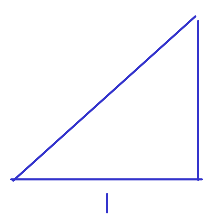
$\int_{-\infty}^{\infty} f(x)\,dx = 1$

$f(x) \geq 0$

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

NO

$\int_{-\infty}^{\infty} f(x)\,dx = = \left(\frac{x^2}{2}\right|_0^1 = \frac{1}{2}$



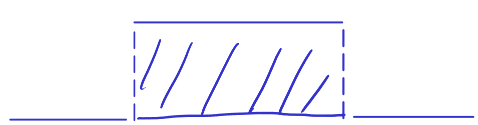
(c) $f(x) = 1$ for $0 \le x \le 1$, and $f(x) = 0$ everywhere else.

YES!

unif $[0, 1]$

$\int_{-\infty}^{\infty} f(x)\,dx = 1$

$f(x) \geq 0$



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

Expo(1) $\quad \exp(-x) = e^{-x}$

YES

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \exp(-x)\,dx = 0-(-1) = 1$$

$$\exp(-x) > 0$$

2. **Disk**

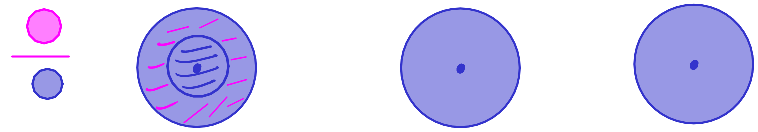
Define a continuous random variable R as follows: we pick a point uniformly at random on a disk of radius 1; the value of R is distance of this point from the center of the disk. We will find the probability density function of this random variable.

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

$\Pr[R \leq \frac{1}{2}] = \frac{1}{4} \neq \frac{1}{2}$

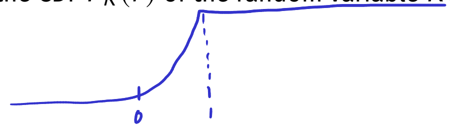
↑CDF of unif[0,1]

$\frac{3}{1} =$



(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?

$$\frac{\pi r^2}{\pi(1)^2} = r^2 \qquad F_R(r) = \begin{cases} r^2 & 0 \leq r \leq 1 \\ 1 & r \geq 1 \\ 0 & r \leq 0 \end{cases}$$



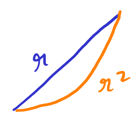
(c) What is the PDF $f_R(r)$ of the random variable R?

$$\frac{d}{dr} F_R(r) = 2r \qquad f_R(r) = \begin{cases} 2r & 0 \leq r \leq 1 \\ 0 & \text{o/w} \end{cases}$$

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

$$F_U(r) = \Pr[U \leq r] \geq \Pr[R \leq r] = F_R(r)$$

$$f_U(r) = 1_{0 \leq r \leq 1} \qquad F_U(r) = \int_{-\infty}^{r} f_U(t)\,dt = \int_0^r 1\,dt = t\Big|_0^r = r$$



$\Pr[U \leq r]$

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).

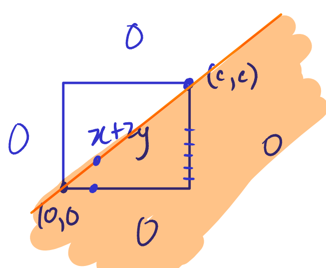
(a) For what value of c is this a valid joint density?

$$\Pr[X = x \cap Y = y] \leftrightarrow f_{X,Y}(x,y)\,dx\,dy$$

$$\iint_{-\infty}^{\infty} f_{X,Y}\,dx\,dy = 1 \qquad f_{X,Y}(x,y) \geq 0$$

$$\iiint f_{X,Y,Z}(x,y,z)\,dx\,dy\,dz = 1$$

$c = \sqrt[3]{\frac{2}{3}}$



$$\int_0^c\int_0^c x + 2y\,dy\,dx = \frac{3c^3}{2} = 1$$

$$\Rightarrow c = \sqrt[3]{\frac{2}{3}}$$

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

$$\Pr[E] = \iint_E f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_0^c\int_0^y x + 2y\,dx\,dy = \int_0^{\sqrt[3]{2/3}} \left[\frac{x^2}{2} + 2yx\right]_0^y dy = \int_0^{\sqrt[3]{2/3}} \frac{5y^2}{2}\,dy = \left(\frac{5}{6}y^3\right|_0^{\sqrt[3]{2/3}} = \frac{5}{9}$$

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

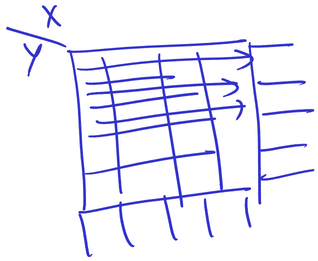
fixed

$$\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,dx = \int_0^{\sqrt[3]{2/3}} x^2 + 2xy\,dx = \left(\frac{x^3}{3} + x^2 y\right|_0^{\sqrt[3]{2/3}} = \frac{2}{9} + \left(\frac{2}{3}\right)^{2/3} y$$

$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{\sqrt[3]{2/3}} x + 2y\,dx = \left(\frac{x^2}{2} + 2yx\right)\Big|_0^{\sqrt[3]{2/3}}$
$= \frac{1}{2}\left(\frac{2}{3}\right)^{2/3} + 2y\left(\frac{2}{3}\right)^{1/3}$

$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$
$= \int_0^c x \frac{f_{X,Y}(x,y)}{f_Y(y)}\,dx$



(d) Compute $E[XY]$. $= \int_0^c \int_0^c \underbrace{xy}_{g(x,y)}(x+2y)\,dx\,dy$

$E[XY] = \underline{E[X]}\,\underline{E[Y]} + \mathrm{Cov}(X,Y)$

LOTUS

$E[g(X,Y)] = \iint g(x,y) f_{X,Y}(x,y)\,dx\,dy$

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.

$X \sim \mathrm{Expo}(\lambda)$ $\quad \underline{Y} = X^\lambda = g(X) \quad g(x) = x^\lambda$

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

$f_X(x) = \lambda \exp(-\lambda x)\,\underline{\mathbb{1}\{x \geq 0\}}$ $= \begin{cases} 1 & x \geq 0 \\ 0 & o/w \end{cases}$

$F_X(x) = \Pr[X \leq x]$

$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$

$= \int_0^x \lambda \exp(-\lambda x) \cdot 1\,dx = \left(-\exp(-\lambda x)\right)\Big|_0^x = -\exp(-\lambda x) - (-1)$
$= \boxed{1 - \exp(-\lambda x)}$

$\int e^{cx}\,dx = \frac{1}{c} e^{cx} + K$

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

$F_Y(y) = \Pr[Y \leq y]$
$= \Pr[X^\lambda \leq y]$
$= \Pr[X \leq y^{1/\lambda}]$
$= F_X(y^{1/\lambda}) = 1 - \exp(-\lambda y^{1/\lambda})$

$X \sim \mathcal{D}$, $g: \mathbb{R} \to \mathbb{R}$, $\frac{dg}{dx} \geq 0$ $\forall x$

$Y = g(X)$, g^{-1} exists.

$f_Y(y)$

$F_Y(y) = \Pr[Y \leq y]$
$= \Pr[g(X) \leq y]$
$= \Pr[X \leq g^{-1}(y)] = F_X(g^{-1}(y))$

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

$\frac{d}{dy} F_Y(y) = \frac{d}{dy}\left(1 - \exp(-\lambda y^{1/\lambda})\right) = \quad y^{\frac{1}{\lambda}-1} e^{-\lambda y^{1/\lambda}}$

$g^{-1}(y) = y^{1/\lambda}$
$g^{-1\prime}(y) = \frac{1}{\lambda} y^{\frac{1}{\lambda}-1}$
$f_X(g^{-1}(y)) = \lambda \exp(-\lambda y^{1/\lambda})$
$\cancel{\frac{1}{\lambda}} y^{\frac{1}{\lambda}-1} \cancel{\lambda} \exp(-\lambda y^{1/\lambda})$ ✓

$f_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = \left(\frac{d}{dy} g^{-1}(y)\right) f_X(g^{-1}(y))$
$= g^{-1\prime}(y) f_X(g^{-1}(y))$

$X \sim \text{Bernoulli}(p)$

$\Pr[X=1] = p$

HHT 1

THTT 2

HHHHHHHT 1

TTTTH 5

$X \sim \text{Binom}(n, p)$ fix

Toss n times, how many H?

$X \sim \text{Geom}(p)$

How many tosses until 1st H? fix

$X \sim \text{Poisson}(\lambda)$

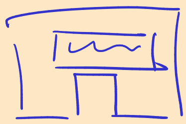
I observe for 1 unit of time. fix

How many arrivals?

$X \sim \text{Expo}(\lambda)$

How long until 1st arrival? fix

3rd ?



λ

"rate"

$\frac{\text{arrivals}}{\text{time}}$

$\lambda = 10 / \text{hr}$

$= 1/6 / \text{min}$

$= 1 / 6\text{mins}$

Normal

Uniform

$\mathcal{N}(0,1)$

$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

$\Phi(x)$

$\mathcal{N}(\mu, \sigma^2)$

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

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?

   (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?

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?

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

   (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?

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:

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

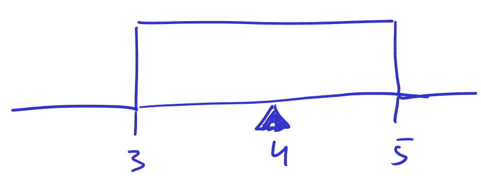
   be the average of the n height samples from the population of penguins, each independently and randomly collected.

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

X_i $\quad \mathbb{E}[X_i] = \int_3^5 \frac{x}{2}\,dx = 4$

$Var(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}^2[X_i] = \frac{49}{3} - 16 = \frac{1}{3}$

$\mathbb{E}[X_i^2] = \int_3^5 x^2 \frac{1}{2}\,dx = \left(\frac{1}{6}x^3\right|_3^5 = \frac{49}{3}$



$E[\hat{X}] = \frac{1}{n} \cdot n \cdot 4 = 4$

$Var(\hat{X}) = \frac{1}{n^2} Var\left(\sum X_i\right) \underset{\text{indep}}{=} \frac{1}{n^2} \sum Var(X_i) \underset{\text{id}}{=} \frac{1}{n^2} \cdot n \cdot \frac{1}{3} = \frac{1}{3n}$

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

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

$\mu \quad c$

95% prob that the interval contains true mean

$\Rightarrow c^2 \geq \frac{20}{3n} \Rightarrow c \geq \sqrt{\frac{20}{3n}}$

$\boxed{\left(4-\sqrt{\frac{20}{3n}},\ 4+\sqrt{\frac{20}{3n}}\right)}$

$|\hat{X}-\mu|>c \quad -c \leq \hat{X}-\mu \leq c \quad \Rightarrow \mu-c \leq \hat{X} \leq \mu+c$

(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 $Pr(-1.96 < \mathcal{N}(0,1) < 1.96) = 0.95$

$\Phi^{-1}(0.025) = 1.96$

$\hat{X} \overset{\text{approx}}{\underset{n\to\infty}{\sim}} \mathcal{N}\left(4, \frac{1}{3n}\right)$

$\frac{\hat{X}-4}{\sqrt{1/3n}} = \sqrt{3n}(\hat{X}-4) \sim \mathcal{N}(0,1)$

$Pr[-c \leq \overbrace{\sqrt{3n}(\hat{X}-4)}^{Z} \leq c] = 0.95$

$X \sim \mathcal{N}(\mu, \sigma^2)$

$\Phi(c) - \Phi(-c) = 2\Phi(c) - 1 = 0.95 \Rightarrow c = 1.96$

$\left(4 - \frac{1.96}{\sqrt{3n}},\ 4 + \frac{1.96}{\sqrt{3n}}\right)$

$\frac{X-\nu}{\tau} \sim \mathcal{N}\left(\mu-\nu, \frac{\sigma^2}{\tau^2}\right)$ $0, 1$

(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?

$\sqrt{20} \approx 4.45$

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 $X > 0.9$.

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

(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.

(d) How do the bounds compare?

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

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

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?)