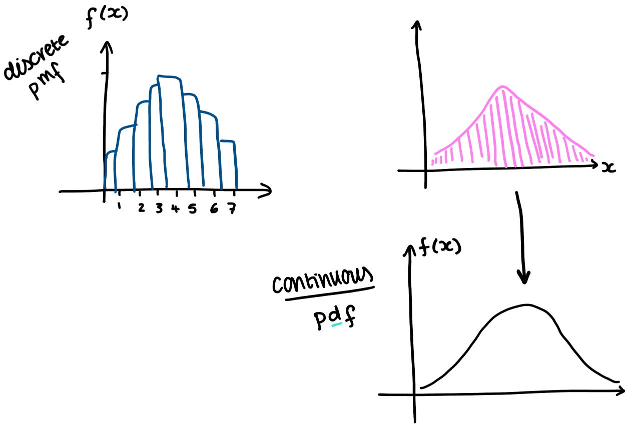
Set 13 and 14 - Continuous Random Variables

February 13, 2023 1:44 PM

Stat 260 Lecture Notes
Sets 13 and 14 - Continuous Random Variables

Recall: A continuous random variable X has an infinite number of possible values and it's impossible to list them all.

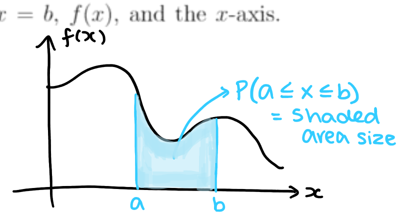
For a discrete random variable we could draw a picture of the pmf $f(x)$ - it looks like a histogram. Imagine making the bars of this histogram thinner and thinner. The top edges of the bars smooth out to a curve - a function. For a continuous random variable X the **probability density function** (pdf) $f(x)$ is this function.



pmf = discrete
pdf = continuous

Rules for the pdf of a continuous random variable X:

- $f(x) \geq 0$ for all x values. → always above the x-axis (discrete version: probabilities ≥ 0)
- The area bounded by the graph of $f(x)$ and the x-axis is 1. That is $\int_{-\infty}^{\infty} f(x)\, dx = 1$. (discrete: pmf probabilities sum to 1)
- $P(a \leq X \leq b)$ = area between $x = a$, $x = b$, $f(x)$, and the x-axis. That is, $P(a \leq X \leq b) = \int_a^b f(x)\, dx$.



1

Rule: If X is a continuous random variable then for a constant c, $P(X = c) = 0$.

so many choices that getting a specific one is unlikely ($\underbrace{\text{so many}}_{}$ infinitely many)

This can be derived from $P(X = c) = P(c \leq X \leq c) = \int_c^c f(x)\, dx = 0$.

Since for a constant c we have that $P(X = c) = 0$, we therefore have that $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.

can drop equality b/c chance of being equal to a or b is 0

Careful! This only applies to continuous random variables. We cannot use this rule if we are working with the binomial distribution or the Poisson distribution (as they are both discrete distributions).

Since $P(X = c) = 0$ for a continuous random variable, when we have a continuous random variable we usually deal with problems like $P(a \leq X \leq b)$ or $P(X \leq a)$ or $P(X \geq a)$.

The **cumulative distribution function** (cdf) $F(x)$ for a continuous random variable is defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$$

dummy variable

where $f(x)$ is the pdf of the random variable X.

random variable is defined by

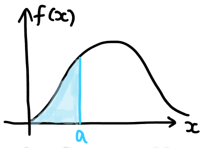
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$$ dummy variable

where $f(x)$ is the pdf of the random variable X.

Say $x = a$. Then $F(a)$ is the area under the pdf curve $f(x)$ to the left of the value $x = a$.

$F(a) = P(X \leq a) = \text{shaded area} = \int_{-\infty}^{a} f(x)\,dx$

cdf



Rule: Suppose X is a continuous random variable with pdf $f(x)$ and cdf $F(x)$. Then at every x where the derivative $F'(x)$ exists, we have that $f(x) = F'(x)$.

pdf is derivative of cdf
cdf is integral of pmf

Even with using calculus, finding areas under the pdf $f(x)$ curve to solve things like $P(X \leq a)$ can be difficult (some integrals may require advanced

techniques such as numerical approximation). In situations like this we often use cdf tables (our stat tables) to look up values for the cdf $F(x) = P(X \leq x)$. If we have knowledge of the exact function $F(x)$ for our cdf, we could also evaluate this function at specific x values to calculate probabilities. (For example, if we wanted to find $P(X \leq 2)$ we could evaluate the function $F(x)$ at $x = 2$, so we could find $F(2)$.)

Note: For a discrete random variable X, where w is next smallest value below a. → often this is just a-1

$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = P(X \leq b) - P(X \leq w)$

not including a

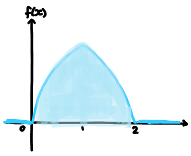
but for a continuous random variable X we have

$P(a \leq X \leq b) = P(X \leq b) - P(X < a)$

$= P(X \leq b) - P(X \leq a)$ → include a

Example 1: Say X is a continuous random variable with pdf

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}.$$



What is the value of c?

We know $1 = \int_{-\infty}^{\infty} f(x)\,dx = \cancel{\int_{-\infty}^{0} 0\,dx}^{\,0} + \int_{0}^{2} c(4x - 2x^2)\,dx + \cancel{\int_{2}^{\infty} 0\,dx}^{\,0}$

$= \int_{0}^{2} c(4x - 2x^2)\,dx = \left[c\left(\frac{4x^2}{2} - \frac{2x^3}{3} \right) \right] \Big|_{0}^{2}$

$= c\left(\frac{4(2^2)}{2} - \frac{2(2^3)}{3} \right) - 0$

$= \frac{8c}{3}$

So $1 = \frac{8c}{3}$

$\Rightarrow c = \frac{3}{8}$.

Find $P(X > 1)$.

$$P(X>1) = P(X \geq 1) = \int_1^{\infty} f(x)\,dx = \int_1^2 \frac{3}{8}(4x - 2x^2)\,dx$$

$$= \left[\frac{3}{8}\left(\frac{4x^2}{2} - \frac{2x^3}{3}\right)\right]_1^2$$

$$= \frac{3}{8}\left(\frac{4(2)^2}{2} - \frac{2(2)^3}{3}\right) - \frac{3}{8}\left(\frac{4(1)^2}{2} - \frac{2(1)^3}{3}\right) = \frac{1}{2}$$

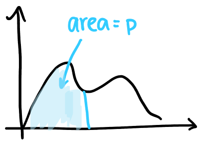
don't have to know how to graph functions

Percentiles: Let p be a value between 0 and 1. The $(100p)^{th}$ **percentile** of the distribution of a continuous random variable X, denoted by $\eta(p)$ is defined by

→ scales decimal

$$p = F(\underbrace{\eta(p)}_{\text{x value}}) = \int_{-\infty}^{\overset{x_0}{\eta(p)}} f(y)\ dy$$

In other words, $\eta(p)$ is the x value where $F(x) = p$, or rather where $P(X \leq x) = p$.

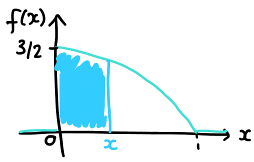


100pth percentile
say it is x_0

Solve for x_0

Example 2: Suppose the pdf of a continuous random variable X is

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$



Solving for cdf

We can find $F(x)$:

dummy variable

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy = \int_0^x \frac{3}{2}(1-y^2)\,dy = \frac{3}{2}y - \frac{3}{2}\frac{y^3}{3}\Big|_0^x$$

$$= \frac{3}{2}x - \frac{x^3}{2}$$

plug in variable

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{3}{2}x - \frac{x^3}{2}, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

because it's a cdf!!!

4

This means that the $(100p)^{th}$ percentile $x = \eta(p)$ satisfies:

x_0

call the percentile $\eta(p) = x_0$

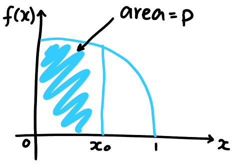
$P(x \le x_0) = p$

$F(x_0) = p$

$\frac{3}{2}x_0 - \frac{x_0^3}{2} = p$

So $0 = \frac{x_0^3}{2} - \frac{3}{2}x_0 + p$

to find percentile, we need to solve this for x_0



For the 50^{th} percentile (that is, when $p = 0.50$) we need to solve:

$0 = \frac{x_0^3}{2} - \frac{3}{2}x_0 + 0.5$

$0 = x_0^3 - 3x_0 + 1$

(multiply both sides by denominator)

This is hard to solve by hand, let's use a computer

$x_0 = \cancel{-1.8794}$

$x_0 = 0.34730$

$x_0 = \cancel{1.5321}$

(x₀ needs to be between 0 and 1)

So, $x_0 = 0.34730$.

$\tilde{\mu} = 0.34730$

This is the 50^{th} percentile.

Note: μ = mean, $\tilde{\mu}$ = median

The **median** $\tilde{\mu}$ is the 50^{th} percentile. (So using the notation, that is that $\eta(0.50) = \tilde{\mu}$.) So half the area under $f(x)$ is to the left of $x = \tilde{\mu}$ and half of the area is to the right.

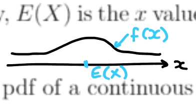
Expected Value and Variance:

For a discrete random variable X: $\mu_X = E(X) = \sum x \cdot f(x) = \sum x \cdot P(X = x)$

integration instead of summation

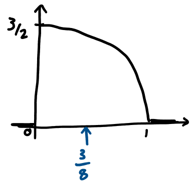
For a continuous random variable X: $\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$.

(Recall: Geometrically, $E(X)$ is the x value that would "balance" the graph of $f(x)$.)



Example 3: Say the pdf of a continuous random variable X is

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$



Find $E(X)$.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$
$$= \int_0^1 x \cdot \tfrac{3}{2}(1 - x^2)\, dx$$
$$= \int_0^1 \tfrac{3}{2}(x - x^3)\, dx$$
$$= \tfrac{3}{2}\left(\tfrac{x^2}{2} - \tfrac{x^4}{4}\right)\Big|_0^1$$
$$= \tfrac{3}{2}\left(\tfrac{1}{2} - \tfrac{1}{4}\right) - 0$$
$$= \tfrac{3}{8}$$

For a discrete random variable X: $E(g(X)) = \sum g(x) \cdot f(x) = \sum g(x) \cdot P(X = x)$

For a continuous random variable X: $E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx$.

Just like before, $E(aX + b) = aE(X) + b$. $\quad E(x^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx$

↳ rules from before hold

6

For a discrete random variable X: $\sigma_X^2 = V(X) = E((X-\mu)^2)$

For a continuous random variable X: $\sigma_X^2 = V(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\, dx$.

The shortcut formula still holds: $V(X) = E(X^2) - [E(X)]^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - \left(\int_{-\infty}^{\infty} x \cdot f(x)\, dx\right)^2$.

The evaluations of $\int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx$ and $\int_{-\infty}^{\infty} x \cdot f(x)\, dx$ are why Math 101 is a corequisite for this course - note that integration by parts may be a useful technique here.

$\int u\,dv = uv - \int v\,du$ — derivative of u

(LIPET)

The standard deviation is still $\sigma_X = \sqrt{V(X)}$.

In example 3

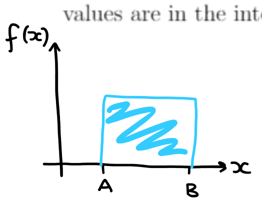
$V(X) = E(X^2) - (E(X))^2 = \int x^2 f(x)\,dx - \left(\int x \cdot f(x)\,dx\right)^2 = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320}$

Q3 on A5 (do on own)

The Uniform Distribution:

X is a uniform random variable if it has pdf $f(x) = c$ where c is a constant.

More specifically, this means that $f(x) = \dfrac{1}{B-A}$ where the possible X values are in the interval $[A, B]$.



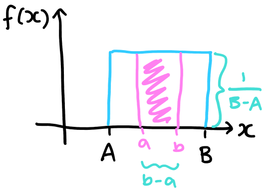
Want total area under curve = 1.

area = (base)(height)

$1 = (B-A)f(x)$

$f(x) = \dfrac{1}{B-A}$

For a uniform random variable X, we have that $P(a \le X \le b) = \frac{b-a}{B-A}$.



$P(a \le x \le b) =$ Shaded area

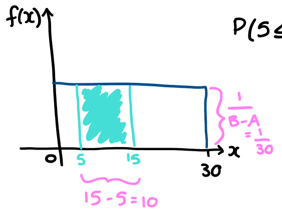
area = (base)(height)

$= (b-a)\left(\frac{1}{B-A}\right)$

$= \frac{b-a}{B-A}$

Example 5: Suppose a person is just as likely (uniform) to arrive at the bus stop any time between 7am and 7:30am. What is the probability that they arrive between 7:05am and 7:15am?

X = # of minutes after 7am when person arrives at bus stop

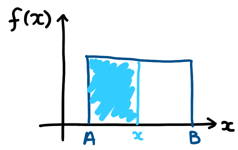


$P(5 \le x \le 15) = \frac{15-5}{30-0} = \frac{10}{30} = \frac{1}{3}$

OR

$P(5 \le X \le 15) =$ shaded area $= 10\left(\frac{1}{30}\right) = \frac{10}{30} = \frac{1}{3}$

Example 6: Let X have a uniform distribution on the interval $[A, B]$. What is the cdf of X? That is, find $F(x)$.



$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$

$= \int_{A}^{x} \frac{1}{B-A}\,dy = \frac{1}{B-A} \cdot y \Big|_{A}^{x}$

$= \frac{1}{B-A} \cdot x - \frac{1}{B-A} \cdot A$

$= \frac{x-A}{B-A}$

$$F(x) = \begin{cases} 0, & x < A \\ \frac{x-A}{B-A}, & A \le x \le B \\ 1, & x > B \end{cases}$$