Probability Theory

Applications for Data Science
Module 4 Continuous Random Variables

Anne Dougherty

March 14, 2021

TABLE OF CONTENTS

Random Variables

At the end of this module, students should be able to

- Define a continuous random variable and give examples of a probability density function and a cumulative distribution function.
- Identify and discuss the properties of a uniform, exponential, and **normal random variable**
- Calculate the expectation and variance of a continuous rv.

In this video we'll review the normal & standard normal distributions, see how they're related to each other, & work through some problems normal distribution

If $X \sim N(\mu, \sigma^2)$ then

density fcn

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \text{ for } -\infty < x < \infty$$

If $Z \sim N(0,1)$ then

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ for } -\infty < x < \infty$$

This means, $P(a \leq Z \leq b) = \int_a^b f_Z(x)\,dx$

Proposition: If $X \sim N(\mu, \sigma^2)$, then $\dfrac{X-\mu}{\sigma} \sim N(0,1)$

Think of $\dfrac{X-\mu}{\sigma}$ as a new rv

It's been **shifted** by μ & scaled by $\frac{1}{\sigma}$

Proposition: If $X \sim N(\mu, \sigma^2)$, then $\underbrace{\frac{X-\mu}{\sigma} \sim N(0,1)}$

this means density fn for $\frac{X-\mu}{\sigma}$ is $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}\left(E(X) - \mu\right) = 0$$

$$V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} V(X) = 1$$

Note: For any cont rv, say Y, with density $f_Y(y)$

we know $P(Y \le a) = \int_{-\infty}^{a} f_Y(y)\,dy$

What if we have $P(2Y \le a) = P(Y \le \frac{a}{2}) = \int_{-\infty}^{a/2} f_Y(y)\,dy$

With $P\Big(\underbrace{\frac{X-\mu}{\sigma}}_{\text{new rv.}} \le a\Big) = P(X \le a\sigma + \mu)$

but we only density fn for X

$$= \int_{-\infty}^{a\sigma+\mu} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}\,dx$$

Do u-subsitution
$u = (x-\mu)/\sigma$
$du = \frac{1}{\sigma}\,dx$

$$= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

Example: If $X \sim N(1, 4)$, (a) find $P(0 \leq X \leq 3.2)$ and find a so that $P(X \leq a) = 0.7$.

(a) $P(0 \leq X \leq 3.2) = P\left(\frac{0-1}{2} \leq \frac{X-1}{2} \leq \frac{3.2-1}{2}\right)$

μ

σ

$= P\left(-\frac{1}{2} \leq \underbrace{\frac{X-1}{2}}_{Z} \leq 1.1\right) = \int_{-.5}^{1.1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$

Recall:

$F_Z(a) = \Phi(a)$

$= P(Z \leq a)$

$= \Phi(1.1) - \Phi(-0.5) \cong .5558$

(b) $P(X \leq a) = P\left(\frac{X-1}{2} \leq \frac{a-1}{2}\right) = P\left(Z \leq \frac{a-1}{2}\right) = \int_{-\infty}^{\frac{a-1}{2}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = .7$

$= \Phi\left(\frac{a-1}{2}\right)$

From Table, $\Phi(.525) = .7$

So $\frac{a-1}{2} = .525$

$\Rightarrow a = 2.05$

move this to next page

Example: The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear-end collisions. Research suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a normal distribution having mean 1.25 seconds and standard deviation 0.46 seconds. What is the probability that the reaction time is between 1 and 1.75 seconds? What assumptions are you making?

X = reaction time, $X \sim N(\mu = 1.25, \sigma^2 = (.46)^2)$

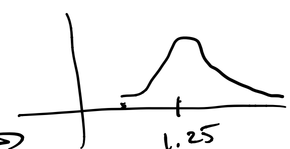
Pause video to work this out on your own

$$P(1 \le X \le 1.75) = P\left(\frac{1-1.25}{.46} \le \frac{X-1.25}{.46} \le \frac{1.75-1.25}{.46}\right)$$

$$= P(-.543 \le Z \le 1.087)$$

$$= \Phi(1.087) - \Phi(-.543)$$

$$\approx .568$$



might be a bit skewed

Assumption: normal rv takes on values from $(-\infty, \infty)$ but most of the prob. is concentrated within 3 std dev. of the mean.

Normal approximation to the binomial distribution

Recall: $X \sim Bin(n, p)$ means that X counts the number of successes in n Bernoulli trials, each with probability of success p. $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \ldots, n$, $E(X) = np$ and $V(X) = np(1-p)$.

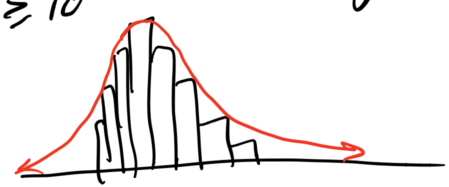
For large n, X can be approximated by a normal rv with $\mu = np$ and $\sigma^2 = np(1-p)$.

location

spread

If $X \sim B(n,p)$ and $np(1-p) \geq 10$ — so n is fairly large

then $\dfrac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$



Example: In a given day, there are approximately 1,000 visitors to a website. Of these, 25% register for a service. Estimate the probability that between 200 and 225 people will register for a service tomorrow.

X = # of people who register for service

$$X \sim B(n = 1000,\ p = .25)$$

$$P(200 \le X \le 225) = \underbrace{\sum_{k=200}^{225} \binom{1000}{k} p^k (1-p)^{1000-k}}_{\text{very difficult to compute directly}}$$

Use $\mu = np = 250$

$\sigma^2 = np(1-p) = 187.5$

$$P\left(\frac{199.5 - 250}{\sqrt{187.5}} \le \frac{X - 250}{\sqrt{187.5}} \le \frac{225.5 - 250}{\sqrt{187.5}} \right)$$

$$= P(-3.688 \le Z \le -1.789) = \Phi(-1.789) - \underbrace{\Phi(-3.688)}_{\approx 0 \text{ to 4 significant digits}}$$

$$= .0367$$