Probability Distributions: Computations with Random Variables

Introduction to Data Science Algorithms
Dirk Grunwald

Overview

- Why?
- Transforming discrete and continuous R.V.'s
- Important case: Normal R.V. $\longrightarrow$ Standard Normal R.V.
- Minimum and Maximum of R.V.'s

Example

A car travels **one mile** at 40m/h and another **mile** at 60m/h.

Total time

$$(\frac{1}{40}h/m + \frac{1}{60}h/m)/2 = \frac{1}{48}h/m$$

or average speed $48m/h$.

Why Compute About Random Variables?

Example

A car travels **one mile** at 40m/h and another **mile** at 60m/h.

Total time

$$(\frac{1}{40}h/m+\frac{1}{60}h/m)/2=\frac{1}{48}h/m$$

or average speed $48m/h$.

Example

A car travels 40 m/h for one **hour**, and 60 m/h for another **hour**.

The mean now is

$$\frac{40\times 1+60\times 1}{2}=50m/h$$

Why Compute About Random Variables?

Example

A car travels **one mile** at 40m/h and another **mile** at 60m/h.

Total time

$$(\frac{1}{40}h/m + \frac{1}{60}h/m)/2 = \frac{1}{48}h/m$$

or average speed $48m/h$.

Example

A car travels 40 m/h for one **hour**, and 60 m/h for another **hour**.

The mean now is

$$\frac{40 \times 1 + 60 \times 1}{2} = 50m/h$$

Wuh?

Computation on Random Variables

- This is an example of

$$E[g(x)] = \sum_i g(x_i) P(X = x_i)$$

.

Computation on Random Variables

- This is an example of

$$E[g(x)] = \sum_i g(x_i) P(X = x_i)$$

.

- In this case

$$g(x) = 1/x$$

Computation on Random Variables

- This is an example of

$$E[g(x)] = \sum_i g(x_i) P(X = x_i)$$

.

- In this case

$$g(x) = 1/x$$

- When we use $g(x) = rx + s$, we have $E[rX + s] = rE[X] + s$, but this isn't true in general.

Jensen's Inequality

- In general

$$g(E[X]) \leq E[g(x)]$$

.

Jensen's Inequality

- In general

$$g(E[X]) \leq E[g(x)]$$

.

- $48 \leq 50$

Jensen's Inequality

- In general
$$g(E[X]) \leq E[g(x)]$$
.
- $48 \leq 50$
- This means that $Var[X] = E[X^2] - (E[X])^2$ will always be positive.

Transforming R.V.'s

Transforming Discrete R.V's

Can usually be done directly on the PMF of the original R.V.

Example

R.V. X represents number of ticket sales for an airplane, assumed to be $X \sim U(0, 200)$.

Transforming Discrete R.V's

Can usually be done directly on the PMF of the original R.V.

Example

R.V. X represents number of ticket sales for an airplane, assumed to be $X \sim U(0, 200)$.

- Plane only holds 150, transform X into R.V. Y that represents number of oversold tickets.

Transforming Discrete R.V's

Can usually be done directly on the PMF of the original R.V.

Example

R.V. X represents number of ticket sales for an airplane, assumed to be $X \sim U(0, 200)$.

- Plane only holds 150, transform X into R.V. Y that represents number of oversold tickets.

$$P(Y = 0) = P(X \leq 150) = \frac{150}{200} = 3/4$$

.

Transforming Discrete R.V's

Can usually be done directly on the PMF of the original R.V.

Example

R.V. X represents number of ticket sales for an airplane, assumed to be $X \sim U(0,200)$.

- Plane only holds 150, transform X into R.V. Y that represents number of oversold tickets.

$$P(Y=0) = P(X \leq 150) = \frac{150}{200} = 3/4$$

.

$$P(Y=k) = P(X=150+k) = 1/200$$

for all $k > 0$.

- In this case, $g(x) = max(x-150,0)$.

Continuous Distributions

- Transformation steps:
  - Transform the Cumulative Distribution Function (CDF), $F_X() \longrightarrow F_Y()$.

Continuous Distributions

- Transformation steps:
  - Transform the Cumulative Distribution Function (CDF), $F_X() \longrightarrow F_Y()$.
  - Take the derivative of $F_Y()$ to produce the PDF.

- Transformation steps:
  - Transform the Cumulative Distribution Function (CDF), $F_X() \longrightarrow F_Y()$.
  - Take the derivative of $F_Y()$ to produce the PDF.

Example

Example: X is R.V. of temperature in Celsius, transform to $Y = \frac{9}{5}X + 32$.

Continuous Distributions

- Transformation steps:
  - Transform the Cumulative Distribution Function (CDF), $F_X() \rightarrow F_Y()$.
  - Take the derivative of $F_Y()$ to produce the PDF.

Example

Example: X is R.V. of temperature in Celsius, transform to $Y = \frac{9}{5}X + 32$.

$$\begin{aligned} F_Y(a) &= P(Y \leq a) = P(\frac{9}{5}X + 32 \leq a) \\ &= P(X \leq \frac{5}{9}(a - 32)) = F_X(\frac{5}{9}(a - 32)) \end{aligned}$$

- Transformation steps:
  - Transform the Cumulative Distribution Function (CDF), $F_X() \rightarrow F_Y()$.
  - Take the derivative of $F_Y()$ to produce the PDF.

Example

Example: X is R.V. of temperature in Celsius, transform to $Y = \frac{9}{5}X + 32$.

$$\begin{aligned} F_Y(a) &= P(Y \leq a) = P(\frac{9}{5}X + 32 \leq a) \\ &= P(X \leq \frac{5}{9}(a-32)) = F_X(\frac{5}{9}(a-32)) \end{aligned}$$

Leads to

$$f_Y(y) = \frac{5}{9} f_X(\frac{5}{9}(y-32))$$

Change of Units Transformations

- IF $Y = rX + s$ for $r, s > 0$, then

Change of Units Transformations

- IF $Y = rX + s$ for $r, s > 0$, then
- CDF transformation:

$$F_Y = F_X(\frac{y-s}{r})$$

- PDF transformation:

$$f_y = \frac{1}{r} f_X(\frac{y-s}{r})$$

Change of Units Transformations

- IF $Y = rX + s$ for $r, s > 0$, then
- CDF transformation:

$$F_Y = F_X(\frac{y-s}{r})$$

- PDF transformation:

$$f_y = \frac{1}{r} f_X(\frac{y-s}{r})$$

- In temperature example, $r = \frac{9}{5}, s = 32$

Change of Units in Normal Distribution

- Let $X \sim N(\mu, \sigma^2)$ then $rX + s \sim N(r\mu + s, r^2\sigma^q)$.

Change of Units in Normal Distribution

- Let $X \sim N(\mu, \sigma^2)$ then $rX + s \sim N(r\mu + s, r^2\sigma^q)$.
- The important case of $r = 1/\sigma, s = -\mu/\sigma$ means

$$X = N(\mu, \sigma) \rightarrow Z = \frac{1}{\sigma}X + \left(-\frac{\mu}{\sigma}\right) = \frac{X - \mu}{\sigma}$$

- and $Z \sim$Standard Normal *e.g.* $N(0, 1)$.

Change of Units in Normal Distribution

- Let $X \sim N(\mu, \sigma^2)$ then $rX + s \sim N(r\mu + s, r^2\sigma^q)$.
- The important case of $r = 1/\sigma, s = -\mu/\sigma$ means

$$X = N(\mu, \sigma) \rightarrow Z = \frac{1}{\sigma}X + \left(-\frac{\mu}{\sigma}\right) = \frac{X - \mu}{\sigma}$$

- and $Z \sim$Standard Normal *e.g.* $N(0, 1)$.
- Historical importance because of standard normal tables, but you'll encounter this transformation over and over in statistics and data science

Example of Standard Normal

Example

Assume $M \sim N(176, 7.47)$ is height of men in north america in cm.

$$Z = \frac{X - 176}{7.47}$$

Example of Standard Normal

Example

Assume $M \sim N(176, 7.47)$ is height of men in north america in cm.

$$Z = \frac{X - 176}{7.47}$$

$$P(X < 190) = P(Z \leq (190 - 176)/7.47) = P(Z \leq 1.87)$$

- Use standard normal table or `stats.norm.cdf(1.87)` 96.7%

Example

Assume $M \sim N(176, 7.47)$ is height of men in north america in cm.

$$Z = \frac{X - 176}{7.47}$$

$$P(X < 190) = P(Z \leq (190 - 176)/7.47) = P(Z \leq 1.87)$$

- Use standard normal table or `stats.norm.cdf(1.87)` 96.7%

99% quantile of $Z = 2.326$. Thus $2.326 * 7.47 + 176$ means 99% of men less than 193.4cm.

Minimum and Maximum

Application

- Given $X_i \sim$ some distribution, what is

$$Z_{max} = max(X_1, X_2, \ldots, X_n)$$

- **1-of-many:** I store n copies of data, each with lifetime X_i. I can recover one copy with distribution Z_{max}.

Application

- Given $X_i \sim$ some distribution, what is

$$Z_{max} = max(X_1, X_2, \ldots, X_n)$$

- **1-of-many:** I store n copies of data, each with lifetime X_i. I can recover one copy with distribution Z_{max}.
- Given $X_i \sim$ some distribution, what is

$$Z_{min} = min(X_1, X_2, \ldots, X_n)$$

- **all-of-many:** I use n computers each with lifetime X_i. Because I need all of them to stay up finish my big data computation, my computation has failure probability Z_{min}.

Derivation

If all X_i have same distribution:

$$\begin{aligned} P(Z_{max} \leq a) &= P(X_1 \leq a, X_2 \leq a, \ldots, X_n \leq a) \\ &= P(X_1 \leq a)P(X_2 \leq a)\ldots P(X_n \leq a) \\ &= F_X(x)^n \end{aligned}$$

It's easy to reduce Z_{min} to the same form if you realize that *any* failure means the same thing as 1-*none failing*.

$$P(Z_{min} \leq a) = 1 - (1 - F(x))^n$$