

AMES Class Notes – Week 7, Wednesday: Probability

Andie Creel

9th October, 2024

1 Random Variables and probability density functions

A **random variable** Y is not inherently random, and while it's called a "variable," it behaves more like a *function* that maps outcomes from a sample space to real numbers (or another set of values). The randomness comes from the fact that the outcome it maps to depends on the result of some random process.

Key Clarifications:

- **Random Variable Y :** This is a function that assigns a number to each possible outcome of a random experiment. It describes a **distribution** over possible values, i.e., the probabilities of different outcomes. The distribution of Y gives us the likelihood of each possible value that Y can take.
- **Specific Value y :** When we observe or "draw" a specific value from this distribution, we denote it as lowercase y . This is a specific realization or outcome of the random variable Y . For example, if Y represents the distribution of grades, then y might represent a specific grade, such as 90%.

Example:

- Y : Represents a random variable describing people's grades, and this random variable is associated with a **distribution** of grades (e.g., a normal distribution of grades in a class).
- y : Represents a specific value that the random variable Y can take, such as a grade of 90%.

Thus, Y describes the entire distribution, and y is one particular outcome from that distribution.

2 Normal Distribution

If the random variable Y is distributed normally it will look like

$$Y = f(y) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy = 1 \quad (1)$$

the inside of the integral is the probability density function. The integral of a pdf must equal 1.

2.1 Defining probability density function (integral of pdf will equal 1)

Consider the random variable Y that has the probability density function (pdf), $f(y)$

$$Y = f(y) = 2ye^{-y^2} \quad (2)$$

and the support is $[0, \infty)$. Definitionally, the area under the pdf will equal 1, because the pdf describes the probability of a random draw of Y equalling y . Because a draw of Y needs to equal something, when you sum up the probabilities of it equally all possible values that needs to sum to 1 (100% chance it does equal something).

Our cumulative distribution function (CDF) is

$$CDF = \int_0^{\infty} 2ye^{-y^2} dy \quad (3)$$

$$= 2 \int_0^{\infty} ye^{-y^2} dy \quad (4)$$

Let $u = -y^2$, $du = -2ydy$

$$= 2 \int_0^\infty ye^u \frac{du}{-2y} \quad (5)$$

$$= -1 \int_0^\infty e^u du \quad (6)$$

$$= -1e^u \Big|_0^\infty \quad (7)$$

$$= -1e^{-y^2} \Big|_0^\infty \quad (8)$$

$$= 0 - -1 \quad (9)$$

$$= 1 \quad (10)$$

Which integrates to 1 after doing a u substitution, and so we know we have a proper pdf. **The integral of a pdf will definitionally be one.**

3 Finding median of a random variable using pdf

What if you want to find the **median**? Set the CDF equal to 0.5 because the median is the 50th percentile.

The unknown is the point on the x-axis (Fig 1) where the area under the pdf is equal to 0.5 (which is defined as our median). Start with equation ?? from our previous integral derivation because we know the integral is equal to that. But replace the upper bound to an unknown \tilde{y} which will be the y value where "half the pdf is to the left" of it and "half is to the right", aka it's our median and we set the CDF from 0 to \tilde{y} equal to 0.5.

$$-1e^{-y^2} \Big|_0^{\tilde{y}} = 0.5 \quad (11)$$

$$-e^{-\tilde{y}^2} - -1 = 0.5 \quad (12)$$

$$-e^{-\tilde{y}^2} = -0.5 \quad (13)$$

$$e^{-\tilde{y}^2} = 0.5 \quad (14)$$

$$\ln(e^{-\tilde{y}^2}) = \ln(1/2) \quad (15)$$

$$-\tilde{y}^2 = \ln(1/2) \quad (16)$$

$$\tilde{y}^2 = -\ln(1/2) \quad (17)$$

$$\tilde{y} = \sqrt{-\ln(1/2)} \quad (18)$$

so we know that the median value of the random variable Y with the pdf in equation ?? is $\tilde{y} = \sqrt{-\ln(1/2)}$.

4 Finding the mean of a random variable using the pdf

To find a mean, we sum the values and divide it by the number of people we summed over. This is the same as multiplying everyone's value by $\frac{1}{N}$,

$$E[Y] = \bar{y} = \frac{1}{N} \sum_n y_n \quad (19)$$

$$= \sum_n \frac{1}{N} y_n \quad (20)$$

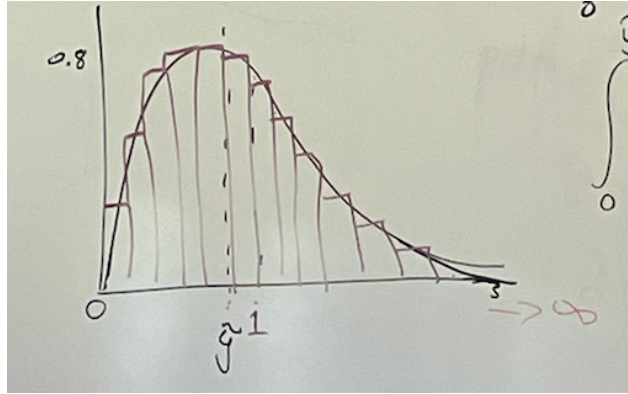


Figure 1: Histogram

where the **expectation function** $E[Y]$ is the same as saying the mean.

We can use the pdf to get our mean! The pdf would replace the $\frac{1}{N}$ and the sum would be replaced by the integral.

$$E[Y] = \bar{y} = \int_0^{\infty} 2ye^{-y^2} * ydy \quad (21)$$

Where the integral sign replaces the summation, and the pdf from equation ?? replaces $1/N$. This is similar to how you may have taken weighted means in the past.

The general rule for finding the mean of a random variable by using the pdf $f(y)$ of that random variable is

$$E[Y] = \int_{\Omega} yf(y)dy \quad (22)$$

where Ω is the support of y (aka any values that the random variable Y can equal).

5 Uniform Distribution Example

Consider the uniform distribution's pdf

$$f(y) = \frac{1}{\theta_2 - \theta_1} \quad (23)$$

Let $\theta_2 = 1$ and $\theta_1 = 0$, which is the same as saying that the random variable Y is only supported over y values of 0 to 1.

$$f(y) = \frac{1}{1 - 0} \quad (24)$$

$$f(y) = 1 \quad (25)$$

Prove $f(y)$ is a **pdf** by showing it integrates to 1.

$$\int_0^1 1dy = y \Big|_0^1 = 1 - 0 = 1 \quad (26)$$

What's the mean?

$$E(Y) = \int_0^1 y * 1 dy = \frac{1}{2} y^2 \Big|_0^1 = \frac{1}{2} \quad (27)$$

What's the median?

$$Med(Y) = \int_0^{\tilde{y}} dy = 0.5 \quad (28)$$

$$y \Big|_0^{\tilde{y}} = 0.5 \quad (29)$$

$$\tilde{y} - 0 = 0.5 \quad (30)$$

6 Finding the variance of a random variable using the pdf

We can find the variance of a random variable Y by considering the squared deviations of a variable y from the mean $E[Y] = \mu$. Our equation for variance is

$$VAR(Y) = E(Y^2) - [E(Y)]^2 \quad (31)$$

We can use the pdf to find the variance of Y , as well, because we can use the pdf to find a mean,

$$VAR(Y) = \int_{\Omega} y^2 f(y) dy - E(Y)^2 \quad (32)$$

6.1 What's the variance of random variable Y with a uniform distribution?

Let's use ?? to find the variance. We already know $E(Y) = 1/2$ from the section above. But how do we get $E(Y^2)$?

We can use the pdf to find the expectation of a *function* of the random variable, which is what $E(Y^2)$ is.

Finding the mean of a *function* of a random variable using the pdf

Let's say you're interested in $g(y)$ rather than y itself.

$$E[g(y)] = \int_{\Omega} g(y) f(y) dy \quad (33)$$

where $f(y)$ continues to be the pdf of Y . This doesn't break Jensen's inequality.

In our current case, $g(Y) = Y^2$

Now we can return to finding the variance. Recall the uniform pdf from equation ?? and that we're assuming the support of random variable Y is 0 to 1, and the $E[Y]$ of the uniformly distributed variable is $1/2$.

$$Var(Y) = \int_0^1 y^2 dy - \frac{1}{2}^2 \quad (34)$$

$$= \frac{1}{3} y^3 \Big|_0^1 - \frac{1}{2}^2 \quad (35)$$

$$= \frac{1}{3} - \frac{1}{4} \quad (36)$$

$$= \frac{1}{12} \quad (37)$$