Proof: PMFs, PDFs, CDFs

Sophie Chowgule, Tom Coleman

Summary

Explanations as to why some PMF's and PDF's are valid.

Before reading this proof sheet, it is recommended that you read Guide: PMFs, PDFs, CDFs. Other recommended reading material will be said when it is needed.

Proof that the binomial distribution is a PMF

Before reading this section, you may find it useful to read [Guide: The binomial theorem].

Remember from Guide: PMFs, PDFs, CDFs that the **binomial distribution** is given by the following.

Binomial distribution

$$P(X = x) = \binom{n}{x} p^x q^{(n-x)} = \frac{n!}{(n-x)!x!} p^x q^{(n-x)}$$

where:

- the random variable $X = x$ measures the number of success in a set of n trials
  - x is number of successes
  - n is number of trials
- p is the probability of success in a single trial
- $q = 1 - p$ is the probability of failure in a single trial

Also from Guide: PMFs, PDFs, CDFs, the two conditions to be a valid PMF are the following:

- **Non-negativity**: The probability assigned to each possible outcome must be greater than or equal to zero, that is:

$$p(x) = P(X = x) \geq 0 \text{ for all values of } x.$$

- **Honesty condition**: The sum of probabilities of all possible outcomes x of a discrete random variable X must be equal to one:

$$\sum_x p(x) = \sum_x P(X = x) = 1.$$

First of all, every term in the PMF for the binomial distribution above is non-negative, and the product of non-negative numbers is non-negative, so $P(X = x) \geq 0$ for any x.

The honesty condition comes about because binomial distributions follow the **binomial theorem**. The binomial theorem states that:

$$\sum_{x=0}^{n} \binom{n}{x} p^x q^{(n-x)} = (p+q)^n$$

(See [Guide: The binomial theorem] for more.)

The number of successes x ranges from 0 (total failure) to n (complete success). Therefore, the sum of all possible probabilities $P(X = x)$ is:

$$\sum_x P(X = x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{(n-x)}$$

which is the left-hand side of the binomial theorem. Using the binomial theorem with $q = 1-p$:

$$\sum_x P(X = x) = (p+q)^n = (p + (1-p))^n = (1)^n = 1$$

So, the sum of the probabilities over all possible values of x equals 1, satisfying the honesty condition.

Proof that the uniform distribution is a PDF

Before reading this section, you may find it useful to read [Guide: Introduction to integration] and [Guide: Properties of integration].

Remember from Guide: PMFs, PDFs, CDFs that the **uniform distribition** over the interval $[a, b]$ is given by the following.

Uniform distribution

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

where a, b are real numbers such that $a < b$.

Also from Guide: PMFs, PDFs, CDFs, the two conditions to be a valid PDF are the following:

- **Non-negativity**: The PDF $f(x)$ must be greater than or equal to zero over its entire range of possible values:

$$f(x) \geq 0 \text{ for all values of } x.$$

- **Honesty condition**: The area under the entire PDF $f(x)$ must be equal to 1, so:

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

To check if this is a valid PDF, you need to confirm that it satisfies these two key conditions.

Non-negativity: $f(x) \geq 0$ for all values of x, as $f(x) = \frac{1}{b-a}$ in $[a, b]$ and 0 otherwise.

Honesty: To satisfy the honesty condition, the integral of the PDF over the interval $[a, b]$ must equal 1. Using the properties of integration, you can split the integral into three parts along the lines of the PDF:

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{a} f(x)\,\mathrm{d}x + \int_{a}^{b} f(x)\,\mathrm{d}x + \int_{b}^{\infty} f(x)\,\mathrm{d}x$$

Using the definition of $f(x)$ on these intervals gives

$$\int_{-\infty}^{a} f(x)\,\mathrm{d}x + \int_{a}^{b} f(x)\,\mathrm{d}x + \int_{b}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{a} 0\,\mathrm{d}x + \int_{a}^{b} \frac{1}{b-a}\,\mathrm{d}x + \int_{b}^{\infty} 0\,\mathrm{d}x$$

Since the integral of 0 over any limits is zero, this reduces to

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 0 + \int_{a}^{b} \frac{1}{b-a}\,\mathrm{d}x + 0 = \int_{a}^{b} \frac{1}{b-a}\,\mathrm{d}x$$

Working out this integral dives

$$\int_a^b \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\int_a^b 1\,\mathrm{d}x = \frac{1}{b-a}\big[x\big]_a^b = \frac{1}{b-a}(b-a) = 1$$

And so you can see that all uniform distributions are valid PDFs.

Proof that the normal distribution is a PDF

Before reading this section, you may find it useful to read [Guide: Properties of integration], [Guide: Integration by substitution], [Guide: Introduction to double integration], and [Guide: Co-ordinate changes in double integration].

Remember from Guide: PMFs, PDFs, CDFs that the **normal distribution** is given by the following.

> **Normal distribution**
>
> $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$
>
> where μ, σ are real numbers such that $\sigma > 0$. (Here, μ is the mean and σ is the standard deviation.)

To check if this is a valid PDF, you need to confirm that it satisfies the two key conditions.

Non-negativity: As an exponential function, $\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) > 0$, and $1/\sigma\sqrt{2\pi} > 0$ as $\sigma > 0$. So $f(x) > 0$.

Honesty: Here's the fun part.

The idea is to show that this integral I, given by

$$I = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)\,\mathrm{d}x$$

is equal to 1. To tackle this integral, it needs to look a little nicer; you can use integration by substitution to do this (see [Guide: Integration by substitution]). Let $u = \frac{x-\mu}{\sigma\sqrt{2}}$. Then $\frac{du}{dx} = \frac{1}{\sigma\sqrt{2}}$, and so $dx = \sigma\sqrt{2}\,du$. As $x \to \pm\infty$, it follows that $u \to \pm\infty$. Since $u^2 = \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$, the integral becomes

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}}e^{-u^2}\,\mathrm{d}u = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-u^2}\,\mathrm{d}u$$

Next, you can use the fact that $\exp(-u^2)$ is an even function to change the limits. Using the property of even function about symmetric limits (see [Guide: Properties of integration]), the integral becomes

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-u^2}\,\mathrm{d}u = \frac{2}{\sqrt{\pi}}\int_{0}^{\infty} e^{-u^2}\,\mathrm{d}u = I$$

All that you have done so far has not changed the value of the integral, so this is still equal to I. Now, the choice of variables in an integral doesn't matter, so $I = \frac{2}{\sqrt{\pi}}\int_0^\infty e^{-v^2}\,\mathrm{d}v$ as well. Multiplying both together gives

$$I^2 = \frac{4}{\pi}\left(\int_0^\infty e^{-u^2}\,\mathrm{d}u\right)\left(\int_0^\infty e^{-v^2}\,\mathrm{d}v\right)$$

Now, the variables here are independent, so you can combine this into a double integral. Doing this gives

$$I^2 = \frac{1}{\pi}\left(\int_0^\infty e^{-u^2}\,\mathrm{d}u\right)\left(\int_0^\infty e^{-v^2}\,\mathrm{d}v\right) = \frac{1}{\pi}\int_0^\infty\int_0^\infty e^{-(u^2+v^2)}\,\mathrm{d}u\,\mathrm{d}v$$

You can now change the co-ordinates to polar co-ordinates (see [Guide: Changing co-ordinates in double integrals] for more). By setting $u = r\cos(\theta)$ and $v = r\sin(\theta)$, it follows that $u^2 + v^2 = r^2$. The region of integration is $0 \le u < \infty$ and $0 \le v < \infty$, which corresponds to the first quadrant of the plane; this is represented in polar co-ordinates by $0 \le r < \infty$ and $0 \le \theta \le \pi/2$. Finally, $\mathrm{d}u\,\mathrm{d}v$ becomes $r\mathrm{d}r\,\mathrm{d}\theta$ by using the Jacobian. Therefore, the integral becomes

$$I^2 = \frac{4}{\pi}\int_0^\infty\int_0^\infty e^{-(u^2+v^2)}\,\mathrm{d}u\,\mathrm{d}v = \frac{4}{\pi}\int_0^{\pi/2}\int_0^\infty re^{-r^2}\,\mathrm{d}r\,\mathrm{d}\theta$$

Now you can evaluate this double integral. The derivative of e^{-r^2} with respect to r is $-2re^{-r^2}$; so that means that the integral of re^{-r^2} is $-\frac{1}{2}e^{-r^2}$ (you can get this result by substitution if you wanted). Using the fact that e^{-r^2} is equal to 1 when $r = 0$ and tends to 0 as r tends to infinity, you can get

$$I^2 = \frac{4}{\pi}\int_0^{\pi/2}\int_0^\infty re^{-r^2}\,\mathrm{d}r\,\mathrm{d}\theta = \frac{4}{\pi}\int_0^{\pi/2}\left[-\frac{1}{2}e^{-r^2}\right]_0^\infty\,\mathrm{d}\theta = \frac{4}{\pi}\int_0^{\pi/2}\frac{1}{2}\,\mathrm{d}\theta$$

Evaluating this final integral gives

$$I^2 = \frac{4}{\pi}\int_0^{\pi/2}\frac{1}{2}\,\mathrm{d}\theta = \frac{4}{\pi}\left[\frac{\theta}{2}\right]_0^{\pi/2} = \frac{4}{\pi}\cdot\frac{\pi}{4} = 1$$

So $I^2 = 1$, implying that $I = \pm 1$. But I cannot be -1, as $f(x)$ is a positive function and the integral of a positive function is always positive. So $I = 1$ and therefore the normal distribution really is a PDF.

Further reading

Guide: PMFs, PDFs, CDFs

Questions: PMFs, PDFs, CDFs

Version history and licensing

v1.0: initial version created in 04/25 by tdhc and Sophie Chowgule as part of a University of St Andrews VIP project.

This work is licensed under CC BY-NC-SA 4.0.