Probability Distributions

Brendon J. Brewer

Department of Statistics
The University of Auckland

https://www.stat.auckland.ac.nz/~brewer/

Suppose a quantity X might be 1, 2, 3, 4, or 5, and we assign probabilities of $\frac{1}{5}$ to each of those possible values. There is some terminology:

Suppose a quantity X might be 1, 2, 3, 4, or 5, and we assign probabilities of $\frac{1}{5}$ to each of those possible values. There is some terminology:

- X is called a 'random variable' (not often by me)

Probability Distributions

Suppose a quantity X might be 1, 2, 3, 4, or 5, and we assign probabilities of $\frac{1}{5}$ to each of those possible values. There is some terminology:

- X is called a 'random variable' (not often by me)
- $\{1, 2, 3, 4, 5\}$ is called the 'sample space', 'hypothesis space', or 'parameter space'

Suppose a quantity X might be 1, 2, 3, 4, or 5, and we assign probabilities of $\frac{1}{5}$ to each of those possible values. There is some terminology:

- X is called a 'random variable' (not often by me)
- $\{1, 2, 3, 4, 5\}$ is called the 'sample space', 'hypothesis space', or 'parameter space'
- $\boldsymbol{p} = \left\{\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right\}$ is the 'probability distribution' for x. In this case, it is a *discrete uniform* distribution.

Probability Distributions

Suppose a quantity X might be 1, 2, 3, 4, or 5, and we assign probabilities of $\frac{1}{5}$ to each of those possible values. There is some terminology:

- X is called a 'random variable' (not often by me)
- $\{1, 2, 3, 4, 5\}$ is called the 'sample space', 'hypothesis space', or 'parameter space'
- $\boldsymbol{p} = \left\{\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right\}$ is the 'probability distribution' for x. In this case, it is a *discrete uniform* distribution.

The probability distribution is often written as $P(X = x) =$ (some function of x).

Properties of discrete probability distributions

Properties of discrete probability distributions

Normalisation:
$\sum_x P(X = x) = 1.$

Properties of discrete probability distributions

Normalisation:
$\sum_x P(X = x) = 1.$

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \sum_x xP(X = x)$

Properties of discrete probability distributions

Normalisation:
$\sum_x P(X = x) = 1.$

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \sum_x xP(X = x)$

Variance:
$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \sum_x (x - \mathbb{E}(X))^2 P(X = x)$

Properties of discrete probability distributions

Normalisation:
$\sum_x P(X = x) = 1.$

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \sum_x xP(X = x)$

Variance:
$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \sum_x (x - \mathbb{E}(X))^2 P(X = x)$

Standard deviation:
$\text{sd}(X) = \sqrt{\text{Var}(X)}$

Properties of discrete probability distributions

Normalisation:
$\sum_x P(X = x) = 1.$

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \sum_x xP(X = x)$

Variance:
$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \sum_x (x - \mathbb{E}(X))^2 P(X = x)$

Standard deviation:
$\text{sd}(X) = \sqrt{\text{Var}(X)}$
The expected value and sd describe the *center* and *width* of the distribution respectively.

Shorthand notation

$P(X = x)$ is cumbersome. x is also just a dummy variable. Common shorthand notation: Use $p(x)$ instead, equivocate between the quantity itself and the dummy variable. E.g.:

$$\mathbb{E}(x) = \sum x p(x) \tag{1}$$

Numerical handling

Numerical handling of discrete probability distributions for a single quantity:

```
xs = np.arange(5, 21) # Grid of possibilities
ps = xs**2            # Not normalised
ps = ps/ps.sum()      # Normalise it
plt.bar(xs, ps)       # Plot it

# Expected value and variance
ex = np.sum(xs*ps)
variance = np.sum(ps*(xs - ex)**2)
np.sum(ps[xs >= 10]) # P(x >= 10)
```

Common discrete distributions

Here are some common discrete distributions:

Common discrete distributions

Here are some common discrete distributions:

- Discrete Uniform (which one is it?)

Common discrete distributions

Here are some common discrete distributions:

- Discrete Uniform (which one is it?)
- Binomial (how many successes out of N quasi-identical trials?)

Common discrete distributions

Here are some common discrete distributions:

- Discrete Uniform (which one is it?)
- Binomial (how many successes out of N quasi-identical trials?)
- Poisson (how many occurrences of rare event?)

Astronomy uses for Discrete Uniform

The number of emission lines in this spectrum is somewhere from 0 to 100, and I don't know the number.

Astronomy uses for Binomial

Suppose it is known (or hypothesised) that 30% of stars of a particular type exhibit a certain kind of oscillation signal. In a new sample of $N = 100$ such stars, let x be the number that have the oscillation.

Then $x \sim \text{Binomial}(100, 0.3)$.

I set a probability equal to a frequency here. What implicit assumption am I making?

Astronomy uses for Poisson

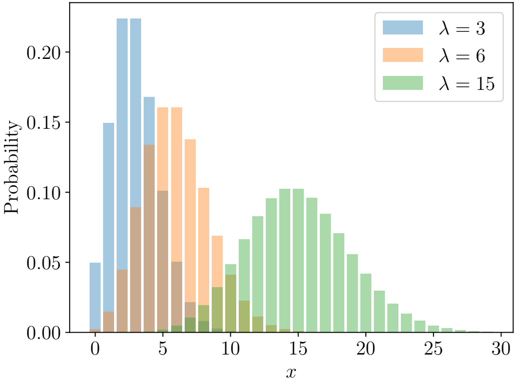
The intensity of an X-ray source is such that you would expect to detect λ photons per minute. Let x be the actual number of photons you observe in a minute.

$$x|\lambda \sim \text{Poisson}(\lambda) \tag{2}$$

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{3}$$

where $\lambda \geq 0$ and $x \in \{0, 1, 2, 3, ...,\}$.

Three Poisson Distributions



Continuous distributions

These are characterised by a continuous hypothesis space (e.g. “all real numbers”) and a *probability density function* (PDF).

For example, normal/gaussian distributions:

$$X|\mu, \sigma \sim \text{Normal}(\mu, \sigma^2) \tag{4}$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \tag{5}$$

$f_X(x)$ is the full notation favoured by many statisticians. You can also just write $f(x)$ or $p(x)$ (and not having any upper-case X).

Properties of continuous probability distributions

Just like discrete, with integrals (implicitly over all x) replacing sums!

Properties of continuous probability distributions

Just like discrete, with integrals (implicitly over all x) replacing sums!
Normalisation:
$\int p(x)\,dx = 1$.

Properties of continuous probability distributions

Just like discrete, with integrals (implicitly over all x) replacing sums!
Normalisation:
$\int p(x)\,dx = 1.$

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \int x p(x)\,dx$

Properties of continuous probability distributions

Just like discrete, with integrals (implicitly over all x) replacing sums!

Normalisation:
$\int p(x)\,dx = 1.$

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \int x p(x)\,dx$

Variance:
$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \int (x - \mathbb{E}(X))^2 f(x)\,dx$

Properties of continuous probability distributions

Just like discrete, with integrals (implicitly over all x) replacing sums!

Normalisation:
$\int p(x)\,dx = 1$.

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \int x p(x)\,dx$

Variance:
$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \int (x - \mathbb{E}(X))^2 f(x)\,dx$

Standard deviation:
$\text{sd}(X) = \sqrt{\text{Var}(X)}$

Properties of continuous probability distributions

Just like discrete, with integrals (implicitly over all x) replacing sums!

Normalisation:
$\int p(x)\,dx = 1$.

Expected value:
$\mathbb{E}(X) = \langle X \rangle = \int xp(x)\,dx$

Variance:
$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \int (x - \mathbb{E}(X))^2 f(x)\,dx$

Standard deviation:
$\text{sd}(X) = \sqrt{\text{Var}(X)}$

The expected value and sd describe the *center* and *width* of the distribution respectively.

Numerical handling of continuous probability distributions for a single quantity:

```
xs = np.linspace(-10.0, 10.0, 10001)  # Grid of possibilities
ps = np.exp(-0.5*((xs - 2.0)/1.5)**2) # Not normalised
ps = ps/np.trapz(ps, x=xs) # Normalise it (integral, not sum)
plt.plot(xs, ps)           # Plot it

# Expected value and variance - integrals, not sums
ex = np.trapz(xs*ps, x=xs)
variance = np.trapz(ps*(xs - ex)**2, x=xs)
np.trapz(ps*(xs >= 5.0), x=xs) # P(x >= 5)
```

Common continuous distributions

Here are some common continuous distributions:

Common continuous distributions

Here are some common continuous distributions:

- Uniform (‘lots of uncertainty’)

Common continuous distributions

Here are some common continuous distributions:

- Uniform (‘lots of uncertainty’)
- Normal/gaussian (commonly used for noise)

Common continuous distributions

Here are some common continuous distributions:

- Uniform ('lots of uncertainty')
- Normal/gaussian (commonly used for noise)
- t (like a normal but with fatter tails)

Common continuous distributions

Here are some common continuous distributions:

- Uniform (‘lots of uncertainty’)
- Normal/gaussian (commonly used for noise)
- t (like a normal but with fatter tails)
- Gamma (nice two-parameter family for a positive quantity)

Common continuous distributions

Here are some common continuous distributions:

- Uniform (‘lots of uncertainty’)
- Normal/gaussian (commonly used for noise)
- t (like a normal but with fatter tails)
- Gamma (nice two-parameter family for a positive quantity)
- Exponential (how long between subsequent events?)

Common continuous distributions

Here are some common continuous distributions:

- Uniform ('lots of uncertainty')
- Normal/gaussian (commonly used for noise)
- *t* (like a normal but with fatter tails)
- Gamma (nice two-parameter family for a positive quantity)
- Exponential (how long between subsequent events?)
- Pareto ('power law')

Astronomy uses for Uniform

There is an asteroid somewhere in an image. Where?

Astronomy uses for Normal/Gaussian

A very common and popular model for how far ‘noise’ will cause a measurement to depart from the true value of what is being measured.

Astronomy uses for t

I recommend using the t-distribution instead of the gaussian distribution for noise if you think a heavier-tailed distribution might be appropriate.

Higher dimensions

Joint probability distributions can be created using the product rule. e.g.,

$$p(x, y) = p(x)p(y|x) \tag{6}$$

The joint distribution allows you to calculate the probability of statements about the *pair* (x, y).

$$P((x, y) \in R) = \int_R p(x, y)\, dx\, dy \tag{7}$$

Marginalisation

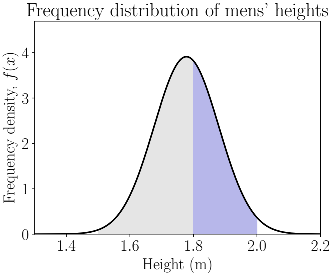
If you have a joint distribution but only care about statements about one of the quantities, you can find the *marginal distribution*:

$$p(x) = \int_{-\infty}^{\infty} p(x, y)\, dy \tag{8}$$

Bayesian and frequentist uses

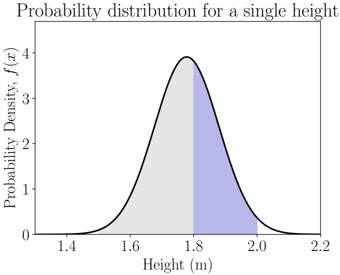
Probability distributions are used in both Bayesian and frequentist senses.

A frequency distribution...



$$
\begin{aligned}
\text{Frac}\left(x \geq 1.8 | x \in [1.5, 2.0]\right) &= \frac{\text{Frac}\left(x \geq 1.8, x \in [1.5, 2.0]\right)}{\text{Frac}\left(x \in [1.5, 2.0]\right)} \\
&= \frac{\int_{1.8}^{2.0} f(x)\,dx}{\int_{1.5}^{2.0} f(x)\,dx}
\end{aligned}
$$

A probability distribution...



$$
\begin{aligned}
\text{Plaus}\left(x \geq 1.8 | x \in [1.5, 2.0]\right) &= \frac{\text{Plaus}\left(x \geq 1.8, x \in [1.5, 2.0]\right)}{\text{Plaus}\left(x \in [1.5, 2.0]\right)} \\
&= \frac{\int_{1.8}^{2.0} f(x)\,dx}{\int_{1.5}^{2.0} f(x)\,dx}
\end{aligned}
$$

Opinion

Important!

The biggest source of confusion in statistics is the failure to distinguish between **frequency distributions** which describe *populations*, and **probability distributions** which describe *uncertainty about a single quantity*[a].

[a] Could be a single non-scalar quantity, such as (3.2, 1.7).

Exercises

See `questions2.pdf`