Continuous Random Variables

PSCI 1801 · Class Handout

Dr. Marc Trussler

This handout is designed as an in-class supplement to the textbook. The material contained on this handout does not represent the full set of information you need to know for the exams. Exclusively studying off of this handout will not be sufficient for the exams.

A continuous *random variable takes on an infinite number of values between two intervals. Because there are infinitely many possible values, the probability of any exact value is 0 — probability is measured over intervals.*

Uniform Random Variables

All points within an interval $[a, b]$ *are equally likely.*

PDF:

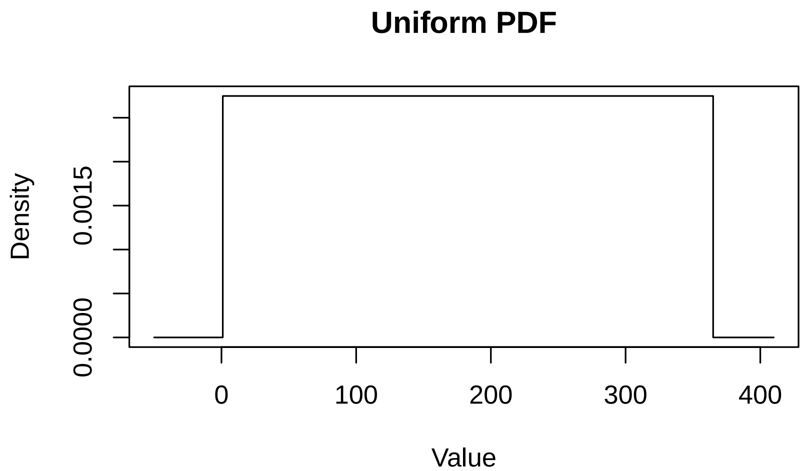
$$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b$$

CDF: the integral of the PDF from $-\infty$ to x. For the uniform, this is a straight line rising from 0 at a to 1 at b.

Expected value and variance:

$$E[X] = \frac{a+b}{2} \qquad V[X] = \frac{1}{12}(b-a)^2$$

R functions: `dunif(x, min, max)`, `punif(x, min, max)`, `runif(n, min, max)`.



The Normal Random Variable

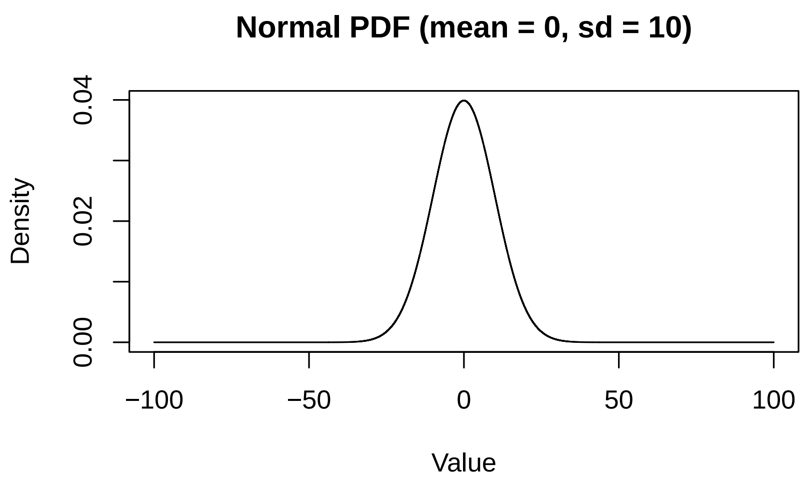
The bell-shaped distribution. Takes any value from $-\infty$ to $+\infty$. Symmetric around its center μ.

PDF (don't memorize; know only that μ and σ fully determine it):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- μ = expected value (also the median because of symmetry)
- σ^2 = variance
- Variance is **independent** of the mean (unlike the binomial)

R functions: `dnorm()`, `pnorm()`, `qnorm()`, `rnorm()`.



Why the normal matters

1. *Many real-world variables are approximately normal (height, IQ, measurement error).*
2. *More importantly for this course: the* sampling distribution *of nearly every statistic we care about (means, proportions, regression coefficients) is normal — regardless of the underlying population. This is what will let us do inference.*

The Standard Normal

Two transformation properties of any normal:

- Shift: if $Z = X + c$, then $Z \sim N(\mu + c, \sigma^2)$
- Scale: if $Z = cX$, then $Z \sim N(c\mu, c^2\sigma^2)$

Combined, these let us convert any *normal into the* standard normal *— the one with $\mu = 0$ and $\sigma = 1$:*

$$Z = \frac{X - \mu}{\sigma}$$

A "z-score" is how many standard deviations away from the mean a value is. All standard-normal probabilities can be looked up in a single table (or from `pnorm(z)` in R).

Empirical rule (approximate):

- 68% of values lie within $\pm 1\sigma$
- 95% within $\pm 2\sigma$ (more precisely, $\pm 1.96\sigma$)
- 99.7% within $\pm 3\sigma$

Answering Questions with These Distributions

Binomial example — left-handed presidents

8 of 45 presidents have been left-handed. 10% of Americans are left-handed. How surprising is this?

P(exactly 8 out of 45):

```
[1] 0.04370462
```

Probability of "at least 8" — use CDF via complement. Note: `1 - pbinom(k, ...)` *gives $P(X > k)$, not $P(X \geq k)$. To get "8 or more," subtract $P(X \leq 7)$:*

```
[1] 0.07569864
```

So about 5% — mildly unusual. Post-WWII (6 of 14 presidents), it's dramatically rarer:

```
[1] 0.001474054
```

Uniform example — birth timing

Uniform on $[0, 365]$. Probability of being born before day 300:

```
[1] 0.8219178
```

Probability of being born in the last 15 days of the year:

```
[1] 0.04109589
```

Probability within $\pm 1\sigma$ of the mean is not *68% here — that only applies to the normal. For this uniform, $\sigma = \sqrt{V[X]} = \sqrt{\frac{365^2}{12}} \approx 105.4$:*

```
[1] 0.5773151
```

Normal example — female height

Female height in US: $\mu = 64$ *inches,* $\sigma = 3$ *inches. Proportion under 5 feet (60 inches):*

```
[1] 0.09121122
```

Two flavors of question to master:

1. **Given a value x, what proportion is to the left/right?** Use `pnorm()`.
2. **Given a probability, what value cuts it off?** Use `qnorm()` (the inverse CDF).

Example — what height defines the 95th percentile?

```
[1] 68.93456
```