Political Science 209 - Fall 2018

Probability III

Florian Hollenbach

8th November 2018

Random Variables and Probability Distributions

- What is a random variable? We assigns a number to an event
  - coin flip: tail= 0; heads= 1
  - Senate election: Ted Cruz= 0; Beto O'Rourke= 1
  - Voting: vote = 1; not vote = 0

Random Variables and Probability Distributions

- What is a random variable? We assigns a number to an event
  - coin flip: tail= 0; heads= 1
  - Senate election: Ted Cruz= 0; Beto O'Rourke= 1
  - Voting: vote = 1; not vote = 0

Probability distribution: Probability of an event that a random variable takes a certain value

Random Variables and Probability Distributions

- P(coin =1); P(coin = 0)
- P(election = 1); P(election = 0)

- Probability density function (PDF): f(x) How likely does X take a particular value?
- Probability mass function (PMF): When X is discrete, $f(x) = P(X = x)$

Random Variables and Probability Distributions

- Probability density function (PDF): $f(x)$ How likely does X take a particular value?
- Probability mass function (PMF): When X is discrete, $f(x) = P(X = x)$
- Cumulative distribution function (CDF): $F(x) = P(X \leq x)$
  - What is the probability that a random variable X takes a value equal to or less than x?
  - Area under the density curve (either we use the sum Σ or integral $\int$)
  - Non-decreasing

Random Variables and Probability Distributions: Binomial Distribution

- PMF: for $x \in \{0, 1, \ldots, n\}$,
  $f(x) \;=\; P(X = x) \;=\; \binom{n}{x} p^x (1-p)^{n-x}$
- PMF function to tell us: what is the probability of x *successes* given n trials with with P(x) = p

Random Variables and Probability Distributions: Binomial Distribution

- PMF: for $x \in \{0, 1, \ldots, n\}$,
  $f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
- PMF function to tell us: what is the probability of x *successes* given n trials with with P(x) = p

In *R*:

```
dbinom(x = 2, size = 4, prob = 0.1) ## prob of 2 successes i
```

```
[1] 0.0486
```

Random Variables and Probability Distributions: Binomial Distribution

- CDF: for $x \in \{0, 1, \ldots, n\}$
  $F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$
- CDF function to tell us: what is the probability of *x or fewer successes* given n trials with with P(x) = p

Random Variables and Probability Distributions: Binomial Distribution

- CDF: for $x \in \{0, 1, \ldots, n\}$
  $F(x) \ = \ P(X \leq x) \ = \ \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$
- CDF function to tell us: what is the probability of *x or fewer successes* given n trials with with P(x) = p

In *R*:

```
pbinom(2, size = 4, prob = 0.1) ## prob of 2 or fewer succes
```

```
[1] 0.9963
```

PMF and CDF

CDF of F(x) is equal to the sum of the results from calculating the PMF for all values smaller and equal to x

PMF and CDF

CDF of F(x) is equal to the sum of the results from calculating the PMF for all values smaller and equal to x

In *R*:

```
pbinom(2, size = 4, prob = 0.1) ## CDF
```

```
sum(dbinom(c(0,1,2),4,0.1)) ## summing up the pdfs
```

```
[1] 0.9963
```

```
[1] 0.9963
```

Random Variables and Probability Distributions: Binomial Distribution

- Example: flip a fair coin 3 times

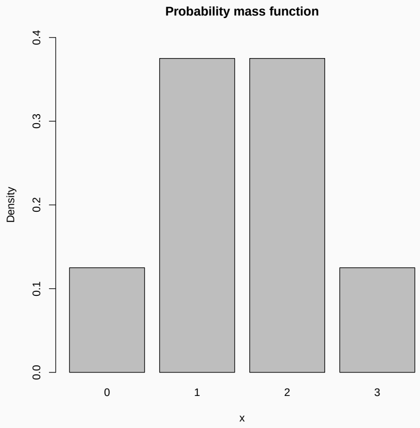
  $f(x) \ = \ P(X = x) \ = \ \binom{n}{x} p^x (1-p)^{n-x}$

  $f(x) \ = \ P(X = 1) \ = \ \binom{3}{1} 0.5^1 (0.5)^2 = 3 * 0.5 * 0.5^2 = 0.375$

Random Variables and Probability Distributions: Binomial Distribution

```
x <- 0:3
barplot(dbinom(x, size = 3, prob = 0.5), ylim = c(0, 0.4), names.arg = x, xlab = "x",
        ylab = "Density", main = "Probability mass function")
```

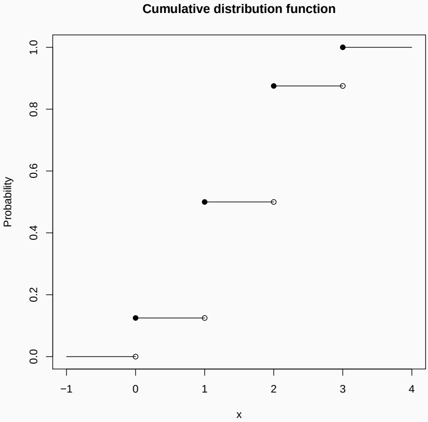
Random Variables and Probability Distributions: Binomial Distribution



Random Variables and Probability Distributions: Binomial Distribution

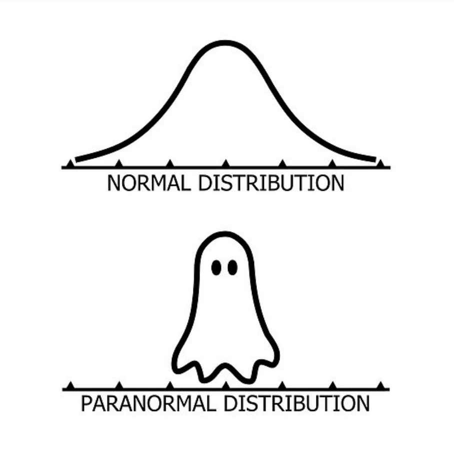
```
x <- -1:4
pb <- pbinom(x, size = 3, prob = 0.5)
plot(x[1:2], rep(pb[1], 2), ylim = c(0, 1), type = "s", xlim = c(-1, 4), xlab = "x",
     ylab = "Probability", main = "Cumulative distribution function")
for (i in 2:(length(x)-1)) {
    lines(x[i:(i+1)], rep(pb[i], 2))
}
points(x[2:(length(x)-1)], pb[2:(length(x)-1)], pch = 19)
points(x[2:(length(x)-1)], pb[1:(length(x)-2)])
```

Random Variables and Probability Distributions: Binomial Distribution



Random Variables and Probability Distributions: Normal Distribution

Normal distribution



Random Variables and Probability Distributions: Normal Distribution

Normal distribution also called Gaussian distribution

Normal distribution

- Takes on values from -∞ to ∞
- Defined by two things: μ and σ^2
    - Mean and Variance (standard deviation squared)

- Mean defines the location of the distribution
- Variance defines the spread

Random Variables and Probability Distributions: Normal Distribution

Normal distribution with mean μ and standard deviation σ

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

Random Variables and Probability Distributions: Normal Distribution

Normal distribution with mean μ and standard deviation σ

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

In *R*:

```
dnorm(2, mean = 2, sd = 2) ## probability of x =2 with norma
```

```
[1] 0.1994711
```

Random Variables and Probability Distributions: Normal Distribution

- CDF (no simple formula. use to compute it):
  $F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$

- What will be F(x =2) for N(2,4)?

Random Variables and Probability Distributions: Normal Distribution

- CDF (no simple formula. use to compute it):
  $F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$
- What will be F(x =2) for N(2,4)?

In *R*:

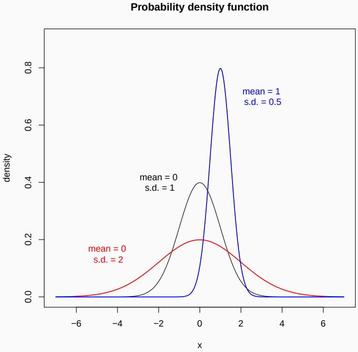
```
pnorm(2, mean = 2, sd = 2) ## probability of x =2 with norma
```

```
[1] 0.5
```

Normal distribution

- Normal distribution is symmetric around the mean
- Mean = Median

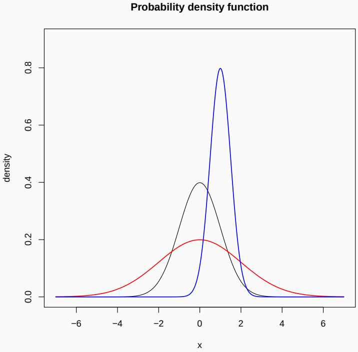
Random Variables and Probability Distributions: Normal Distribution



Random Variables and Probability Distributions: Normal Distribution in R

```
x <- seq(from = -7, to = 7, by = 0.01)
plot(x, dnorm(x), xlab = "x", ylab = "density", type = "l",
     main = "Probability density function", ylim = c(0, 0.9))
lines(x, dnorm(x, sd = 2), col = "red", lwd = lwd)
lines(x, dnorm(x, mean = 1, sd = 0.5), col = "blue", lwd = lwd)
```

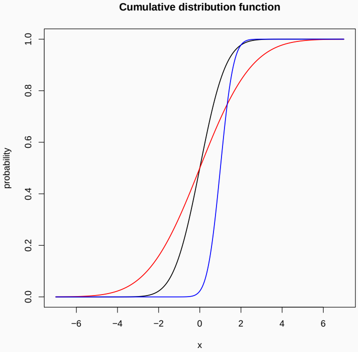
Random Variables and Probability Distributions: Normal Distribution in R



Random Variables and Probability Distributions: Normal Distribution in R

```
plot(x, pnorm(x), xlab = "x", ylab = "probability", type = "l",
     main = "Cumulative distribution function", lwd = lwd)
lines(x, pnorm(x, sd = 2), col = "red", lwd = lwd)
lines(x, pnorm(x, mean = 1, sd = 0.5), col = "blue", lwd = lwd)
```

Random Variables and Probability Distributions: Normal Distribution in R



Random Variables and Probability Distributions: Normal Distribution

Let $X \sim N(\mu, \sigma^2)$, and c be some constant

- Adding/subtracting to/from a random variable that is normally distributed also results in a variable with a normal distribution: $\mathsf{Z} = \mathsf{X} + \mathsf{c}$ then $Z \sim N(\mu + c, \sigma^2)$

Random Variables and Probability Distributions: Normal Distribution

Let $X \sim N(\mu, \sigma^2)$, and c be some constant

- Adding/subtracting to/from a random variable that is normally distributed also results in a variable with a normal distribution: Z = X + c then $Z \sim N(\mu + c, \sigma^2)$
- Multiplying or dividing a random variable that is normally distributed also results in a variable with a normal distribution: $Z = X \times c$ then $Z \sim N(\mu \times c, (\sigma \times c)^2)$
- Z-score of a random variable that is normally distributed has mean 0 and sd = 1

Random Variables and Probability Distributions: Normal Distribution

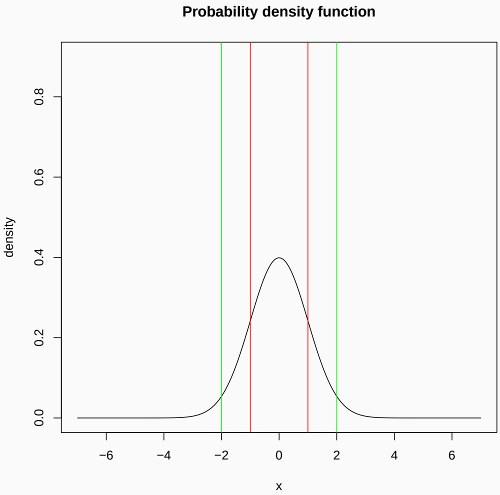
Curve of the standard normal distribution:

- Symmetric around 0
- Total area under the curve is 100%
- Area between -1 and 1 is ~68%
- Area between -2 and 2 is ~95%
- Area between -3 and 3 is ~99.7%

Random Variables and Probability Distributions: Normal Distribution

```
x <- seq(from = -7, to = 7, by = 0.01)
lwd <- 1.5
plot(x, dnorm(x), xlab = "x", ylab = "density", type = "l",
     main = "Probability density function", ylim = c(0, 0.9))
abline(v= -1, col = "red")
abline(v= 1, col = "red")
abline(v= -2, col = "green")
abline(v= 2, col = "green")
```

Random Variables and Probability Distributions: Normal Distribution



Random Variables and Probability Distributions: Normal Distribution

Curve of the any normal distribution:

- Symmetric around 0
- Total area under the curve is 100%
- Area between -1SD and +1SD is ~68%
- Area between -2SD and +2SD is ~95%
- Area between -3SD and +3SD is ~99.7%

Expectations, *Means*, and *Variances*

For probability distributions, means should not be confused with *sample means*

Expectations or means of a random variable have specific meanings for its the probability distribution

Means and Expectation

A sample mean varies from sample to sample

Mean of a probability distribution is a theoretical construct and constant

Means and Expectation

A sample mean varies from sample to sample

Mean of a probability distribution is a theoretical construct and constant

Example: Age of undergraduate body at A&M

The expectation of a random variable is equal to the sum of all possibilities *weighted* by the probabilities

Means and Expectation

The expectation of a random variable is equal to the sum of all possibilities *weighted* by the probabilities

Example: expectation of rolling one die

$\mathbb{E}(X) = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4\frac{1}{6} \times 5\frac{1}{6} \times 6 = 3.5$

Means and Expectation

The expectation of a random variable is equal to the sum of all possibilities *weighted* by the probabilities

$$\mathbb{E}(X) = \begin{cases} \sum_x x \ f(x) & \text{if } X \text{ is discrete} \\ \int x \ f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Means and Expectation

Remember the lottery!

Expected value: winnings $\times$ p(winning) $+ 0 \times$ p(not winning)

Means and Expectation

What is $\mathbb{E}(X)$ for the number of heads in 100 coin flips?

Means and Expectation

What is $\mathbb{E}(X)$ for the number of heads in 100 coin flips?

$\mathbb{E}(X) = 0.5 \times 1 + 0.5 \times 1 + ... + 0.5 \times 1 = 0.5 * 100 = 50$

Variance

- Variance is standard deviation squared
- Variance in a probability distribution indicates how much uncertainty exists
- Similar but not the same as sample standard deviation

Variance

Population variance:

$$\mathbb{V}(X) \;=\; \mathbb{E}[\{X - \mathbb{E}(X)\}^2] \;=\; \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2$$