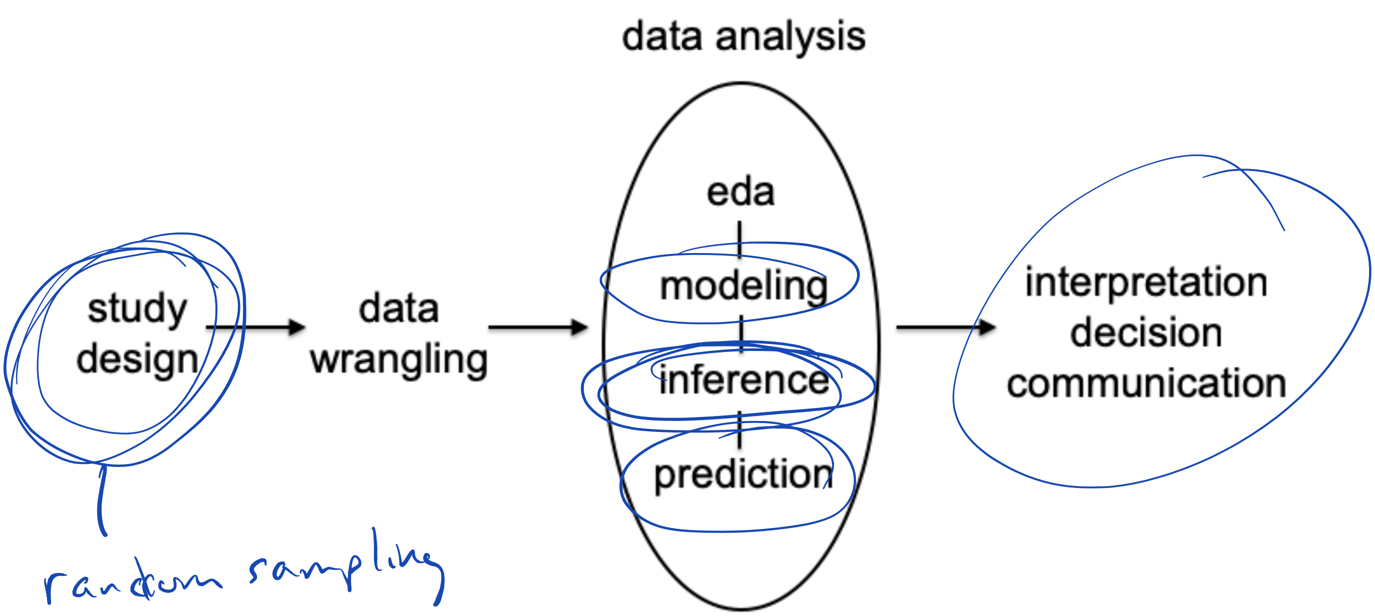
Random Variables

- Probability
- Random variable
- Distributions of rv's
- Convergence of rv's
- Likelihood
- EFDs

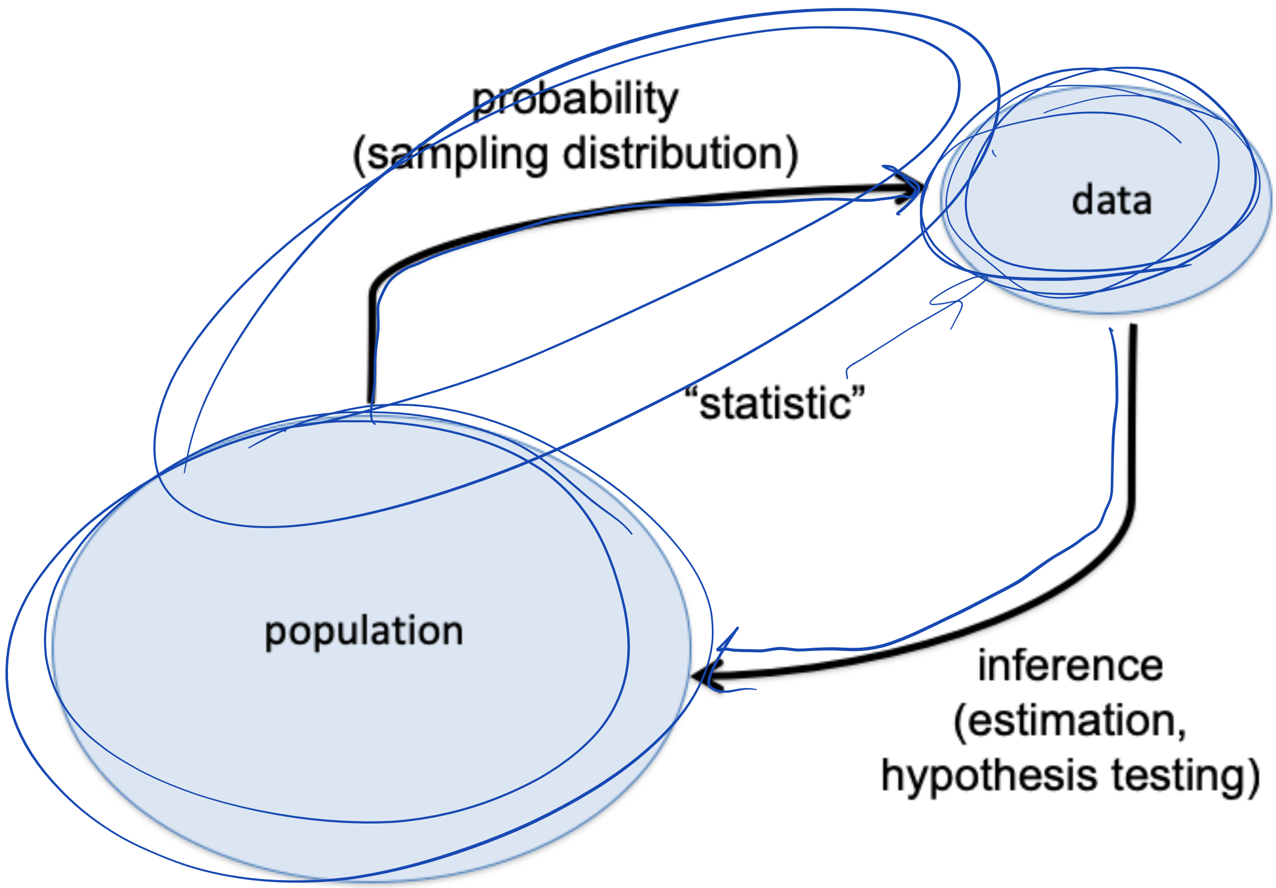


random sampling
randomization

Doesn't use probability:

- descriptive statistics
- exploratory data analysis
- sometimes prediction

Central Dogma of Statistical Inference



Probability Theory

- Sample space: Call it Ω, the set of outcomes
- Interested in subsets $A \subseteq \Omega$ and calculating the probability of A, called events

- Examples:

  - $\Omega = \{HH, HT, TH, TT\}$ coin flips
  - $\Omega = \{AA, AT, TT\}$ genotypes
  - $\Omega = [0, \infty)$ heights

Measure Theoretic Probability

Williams

$$(\Omega, \mathcal{F}, Pr)$$

- Ω sample space
- $\mathcal{F}$ σ-algebra of subsets A where probability
- Pr is the probability measure

Mathematical Probability

1. The probability of any event A is such that
   $0 \le Pr(A) \le 1$

2. If Ω is the sample space, then

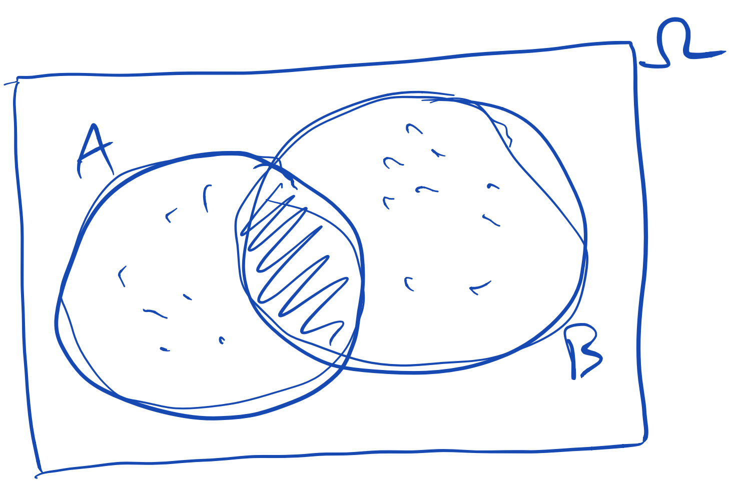
$Pr(\Omega) = 1$

3. Let A^c be the outcomes from Ω not in A, then $Pr(A) + Pr(A^c) = 1$

4. For any n events $A_1, A_2, \dots, A_n$ such that $A_i \cap A_j = \phi$ for all $i \neq j$ then $Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} Pr(A_i)$

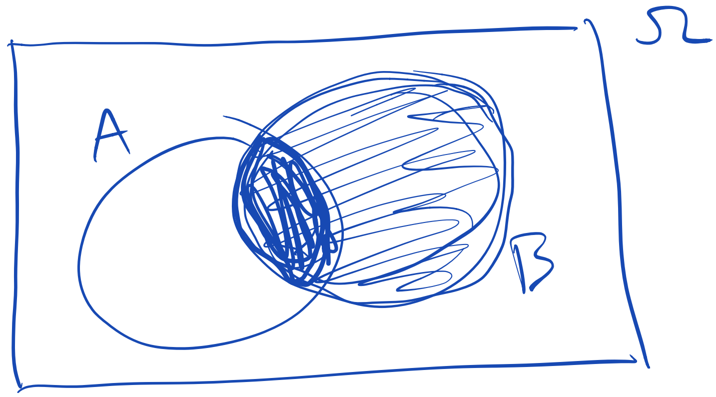
Union of two events

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$



Conditional Probability

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} \Rightarrow \begin{array}{l} Pr(A \cap B) \\ = Pr(A|B)Pr(B) \end{array}$$



Independence

Two events A and B are "independent" when:

(1) $Pr(A|B) = Pr(A)$

(2) $Pr(B|A) = Pr(B)$

(3) $Pr(A \cap B) = Pr(A)Pr(B)$

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = Pr(A)$$

$$\Rightarrow Pr(A \cap B) = Pr(A)Pr(B)$$

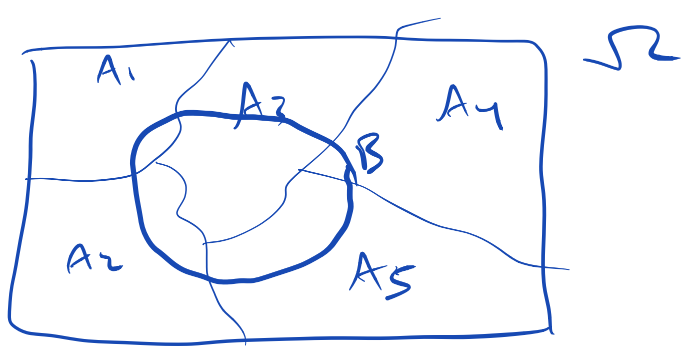
Bayes Theorem

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$$

Law of total probability

Events $A_1, A_2, \dots, A_n$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^{n} A_i = \Omega$, it follows for event B, $\sum_{i=1}^{n} \Pr(B \cap A_i)$

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B|A_i)\Pr(A_i)$$



Random Variable

A random variable X is a function from Ω to $\mathbb{R}$:

$$X : \Omega \longrightarrow \mathbb{R}$$

For any outcome $\omega \in \Omega$, the function $X(\omega)$ produces a real number

The range of X is

$$\mathcal{R} = \{ X(\omega) : \omega \in \Omega \}$$

$$\mathcal{R} \subseteq \mathbb{R}$$

Distribution of RVs

1. Probability mass function (pmf) or probability density function (pdf)

2. cumulative distribution function (cdf)

Discrete RVs

where $\mathcal{R}$ is discrete, e.g.,

$\mathcal{R} = \{0, 1, 2, 3, \ldots\}$

Use pmf:

$$f(x) = \Pr(X = x) \quad \text{for } x \in \mathcal{R}$$

The cdf:

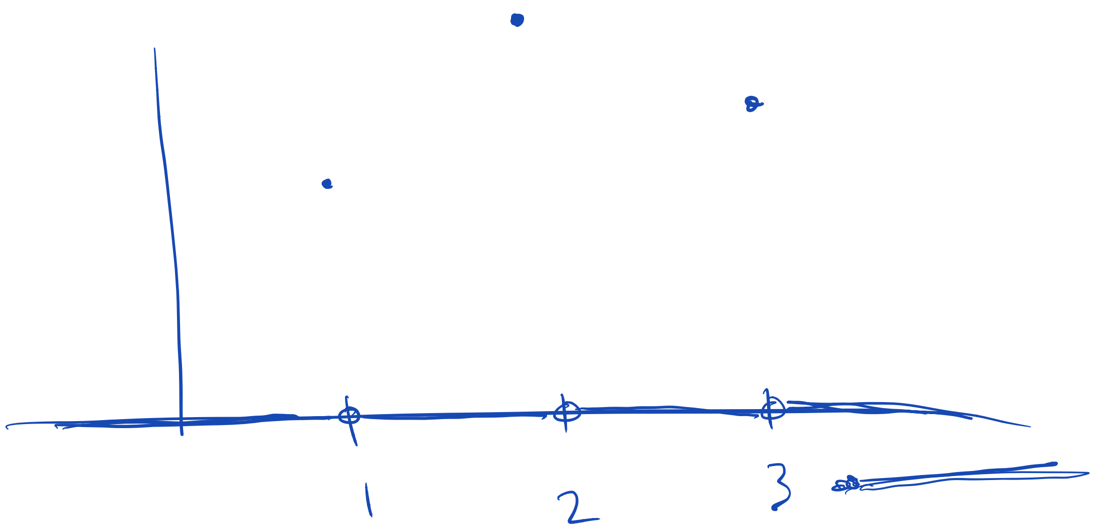
$$F(y) = \Pr(X \leq y) = \sum_{x \leq y} f(x)$$

$$= \sum_{x \leq y} \Pr(X = x)$$

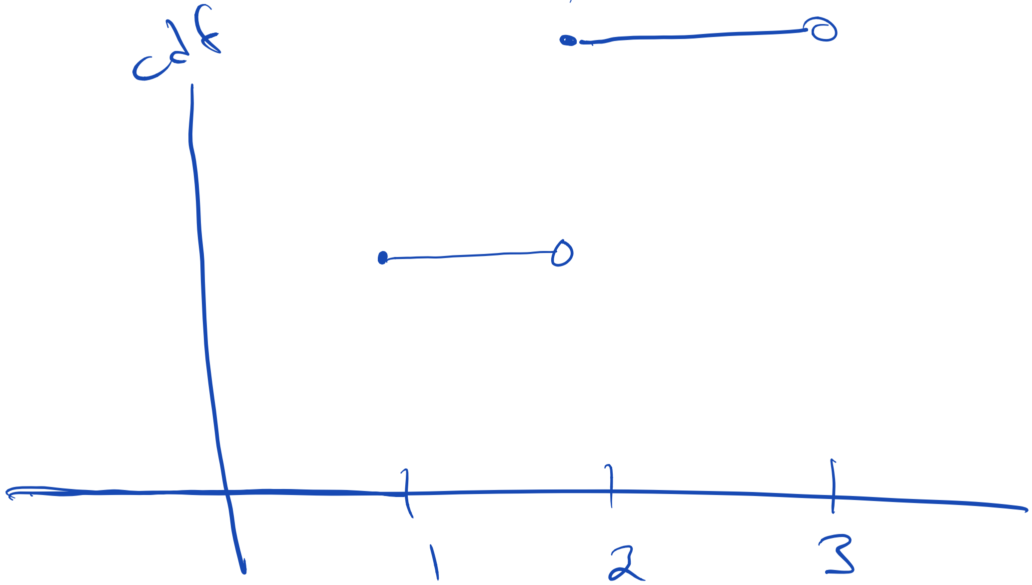
for $y \in \mathbb{R}$

$$\sum_{x \in \mathcal{R}} f(x) = 1 \quad \text{since } \Pr(\Omega) = 1$$

pmf:



cdf



$$\Pr(X \le b) = F(b) = \sum_{x \le b} f(x)$$

$$\Pr(X \ge a) = 1 - \Pr(X < a)$$
$$= 1 - \Pr(X \le a-1)$$
$$= 1 - F(a-1)$$

Continuous RVs

where $\mathcal{R}$ is a continuous interval

E.g. $\mathcal{R} = [0,1]$

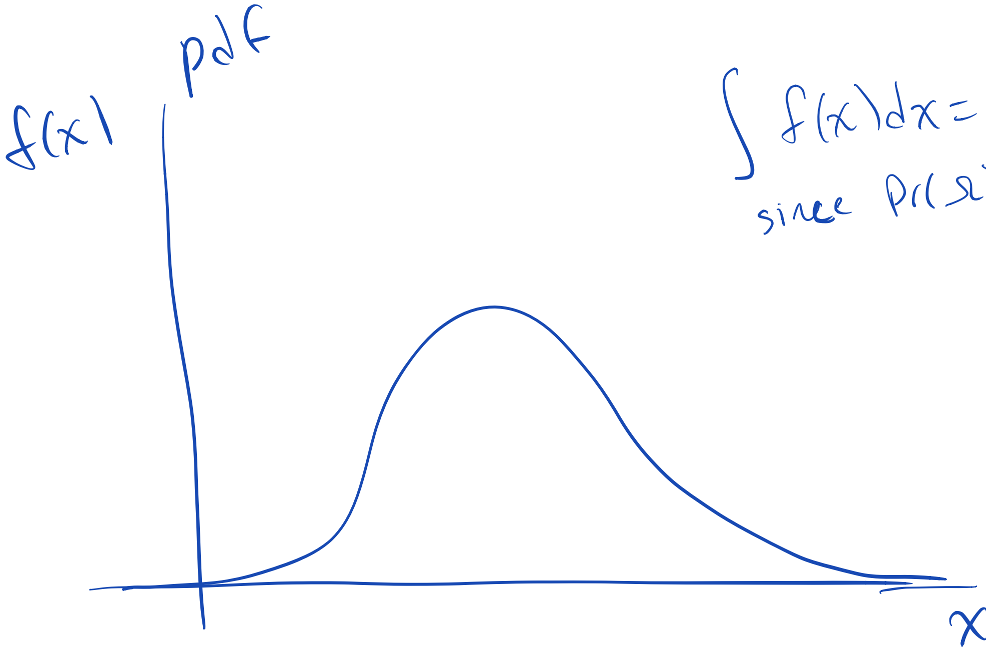
$\mathcal{R} = \mathbb{R}$

The pdf

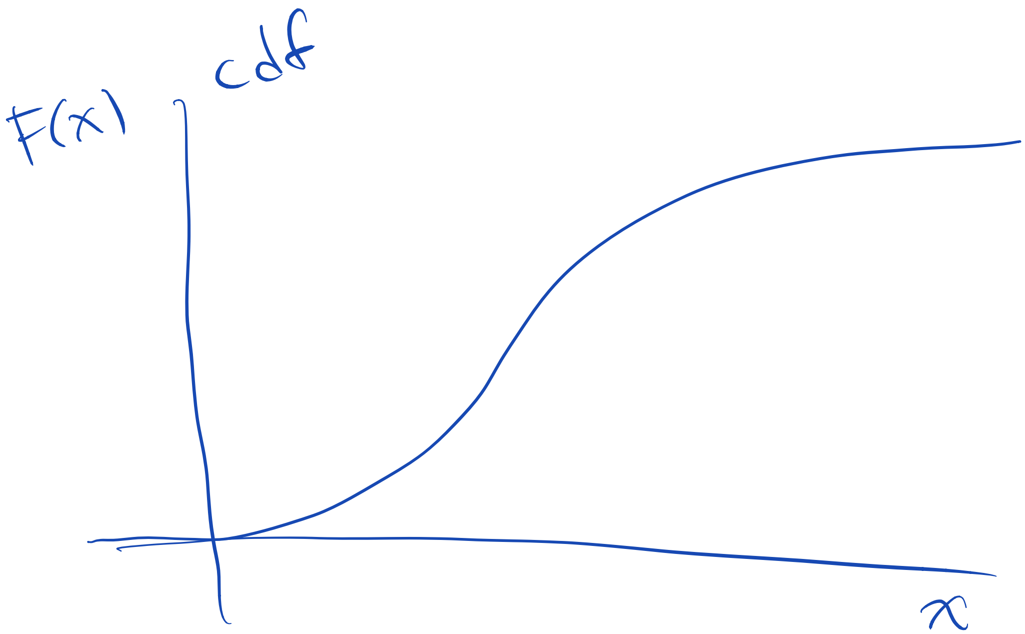
$f(x)$ the "probability" of being in an infinitesimal interval of x

The cdf

$$F(y) = \Pr(X \le y) = \int_{-\infty}^{y} f(x)\,dx$$



$\int f(x)\,dx = 1$
since $Pr(\Omega) = 1$



cdf

(1) $F(y) = \sum_{x \le y} f(x)$ discrete

(2) $F(y) = \int_{-\infty}^{y} f(x)\,dx$ continuous

(3) $F(y) = \int_{-\infty}^{y} dF(x)$ measure theory

CDFs

- they are right continuous
- $F(y) = 0$ as $y \to -\infty$
- $F(y) = 1$ as $y \to \infty$
- the right derivative of $F(y)$ at x is $f(x)$

Sample vs. Population Statistics

"sample statistics" are calculations on observed data

"population statistics" are calculations on the entire population, i.e., on the probability distribution

Expected Value (population mean)

$$E[X] = \sum_{x \in \mathbb{R}} x \Pr(X = x) = \sum_{x \in \mathbb{R}} x f(x)$$

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$E[X] = \int_{-\infty}^{\infty} x\, dF(x)$$

Population Variance

$$Var(X) = E[(X - E[X])^2]$$

$$= \sum_{x \in R} (x - E[X])^2 f(x)$$

$$\left\{ \frac{\sum (x_i - \bar{x})^2}{n-1} \right\}$$

$$= E[X^2] - E[X]^2$$

$$E[X^2] = \sum_{x \in R} x^2 f(x)$$

$$E[h(X)] = \sum_{x \in R} h(x) f(x)$$

$$\mathrm{Var}(X) = E[(X - E(X))^2]$$
$$= \int [x - E[X]]^2 f(x)\,dx$$

Covariance & Correlation

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

$$SD(X) = \sqrt{\mathrm{Var}(X)}$$

Discrete RVs

Uniform (Discrete)

This simple rv distribution assigns equal probabilities to a finite set of values:

$$X \sim \text{Uniform}\{1, 2, \ldots, n\}$$

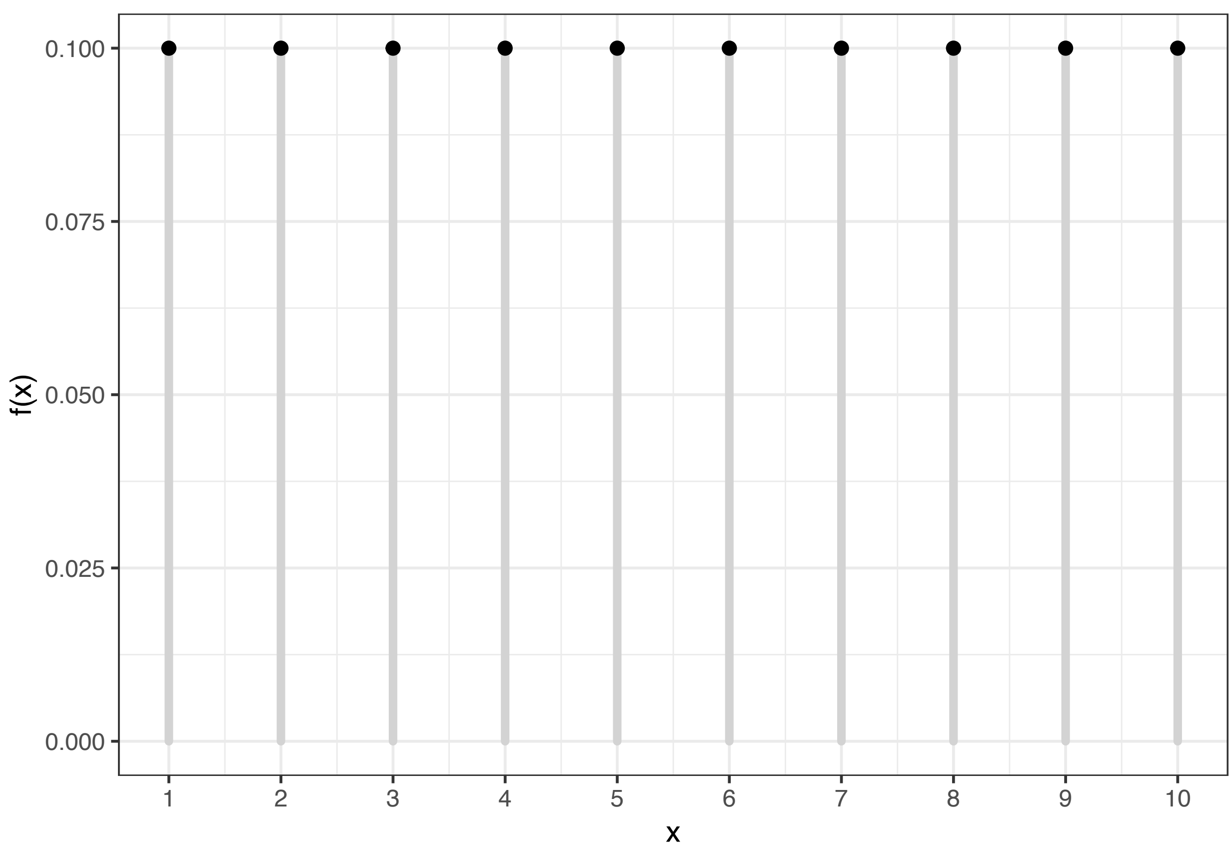
$$\mathcal{R} = \{1, 2, \ldots, n\}$$

$$f(x; n) = 1/n \text{ for } x \in \mathcal{R}$$

$$\text{E}[X] = \frac{n+1}{2}, \;\; \text{Var}(X) = \frac{n^2 - 1}{12}$$

Uniform (Discrete) PMF

$n=10$



Uniform (Discrete) in R

There is no family of functions built into R for this distribution since it is so simple. However, it is possible to generate random values via the `sample` function:

```
> n <- 20L
> sample(x=1:n, size=10, replace=TRUE)
 [1]  8 19  4  1 18 15 18 18  2  7
>
> x <- sample(x=1:n, size=1e6, replace=TRUE)
> mean(x) - (n+1)/2
[1] 0.006991
> var(x) - (n^2-1)/12
[1] 0.0208284
```

Bernoulli

A single success/failure event, such as heads/tails when flipping a coin or survival/death.

$$X \sim \text{Bernoulli}(p)$$

parameter

$$\mathcal{R} = \{0, 1\}$$

$$f(x; p) = p^x (1-p)^{1-x} \text{ for } x \in \mathcal{R}$$

$f(0) = 1-p$

$f(1) = p$

$$\mathrm{E}[X] = p, \ \mathrm{Var}(X) = p(1-p)$$

Binomial

An extension of the Bernoulli distribution to simultaneously considering n independent success/failure trials and counting the number of successes.

$$X \sim \text{Binomial}(n, p)$$

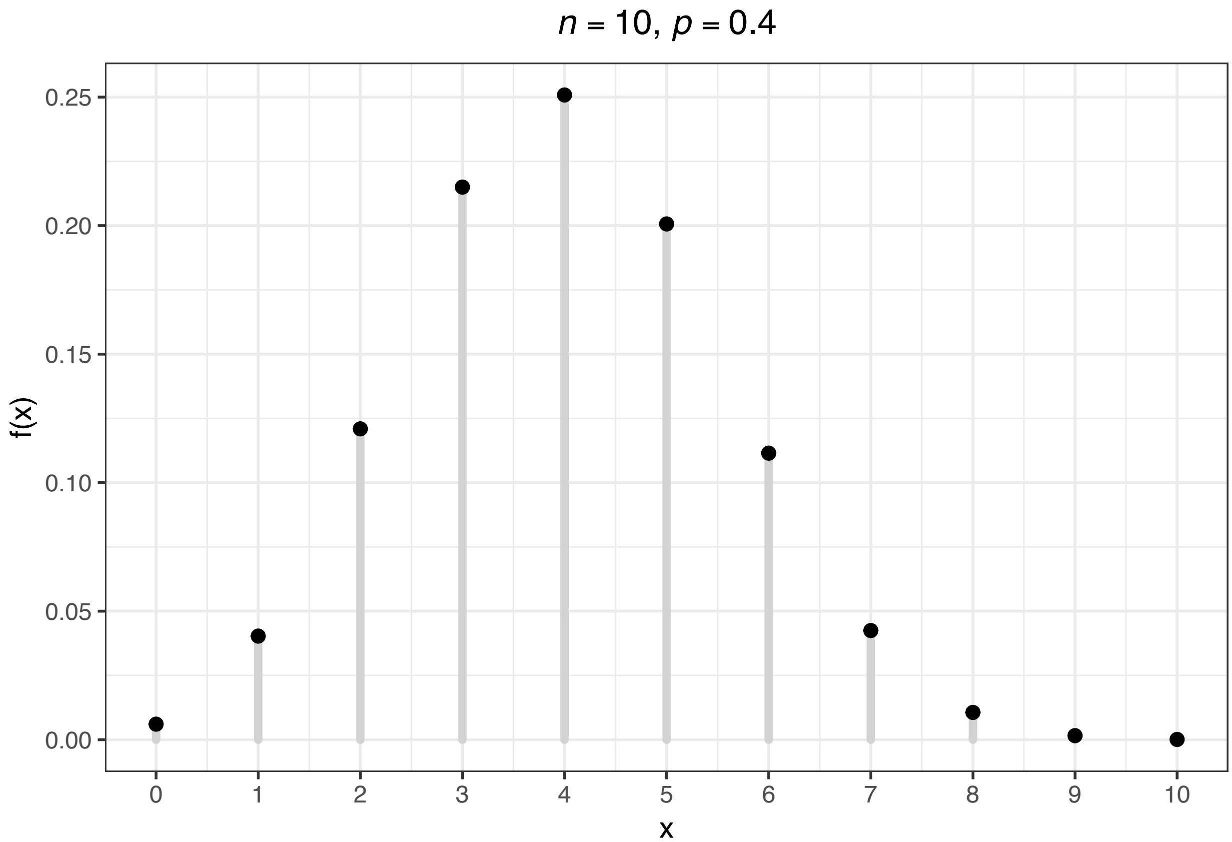
$$\mathcal{R} = \{0, 1, 2, \ldots, n\}$$

$$f(x; p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x \in \mathcal{R}$$

$$\mathrm{E}[X] = np, \ \mathrm{Var}(X) = np(1-p)$$

Note that $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the number of unique ways to choose x items from n without respect to order.

Binomial PMF



Binomial in R

pmf

cdf

$F(y) = P(X \le y)$

$Pr(X \ge y)$

```
> str(dbinom)
function (x, size, prob, log = FALSE)
```

```
> str(pbinom)
function (q, size, prob, lower.tail = TRUE, log.p = FALSE)
```

```
> str(qbinom)
function (p, size, prob, lower.tail = TRUE, log.p = FALSE)
```

```
> str(rbinom)
function (n, size, prob)
```

Poisson

Models the number of occurrences of something within a defined time/space period, where the occurrences are independent. Examples: the number of lightning strikes on campus in a given year; the number of emails received on a given day.

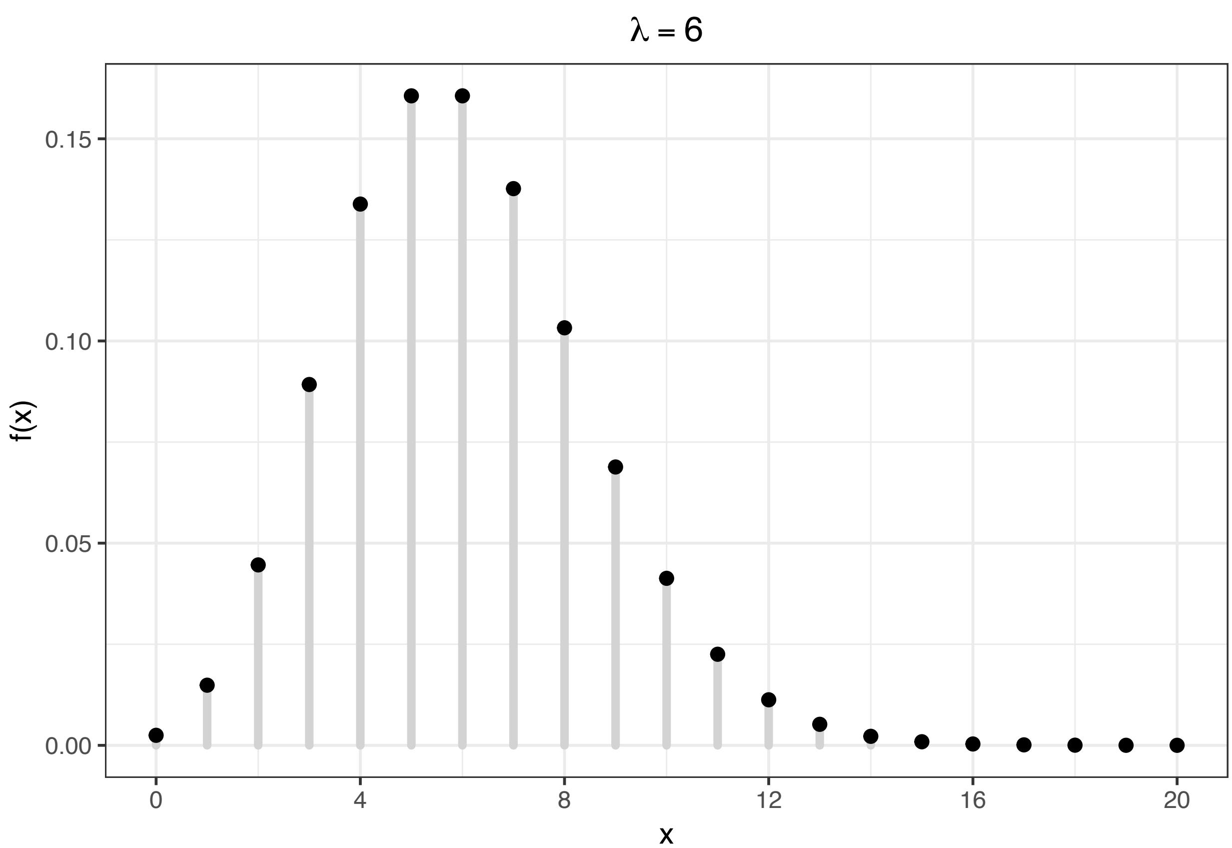
$$X \sim \text{Poisson}(\lambda)$$

$$\mathcal{R} = \{0, 1, 2, 3, \ldots\}$$

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x \in \mathcal{R}$$

$$\text{E}[X] = \lambda, \ \text{Var}(X) = \lambda$$

Poisson PMF



Poisson in R

```
> str(dpois)
function (x, lambda, log = FALSE)
```

```
> str(ppois)
function (q, lambda, lower.tail = TRUE, log.p = FALSE)
```

```
> str(qpois)
function (p, lambda, lower.tail = TRUE, log.p = FALSE)
```

```
> str(rpois)
function (n, lambda)
```

Continuous RVs

Uniform (Continuous)

Models the scenario where all values in the unit interval [0,1] are equally likely.

$$X \sim \text{Uniform}(0, 1)$$

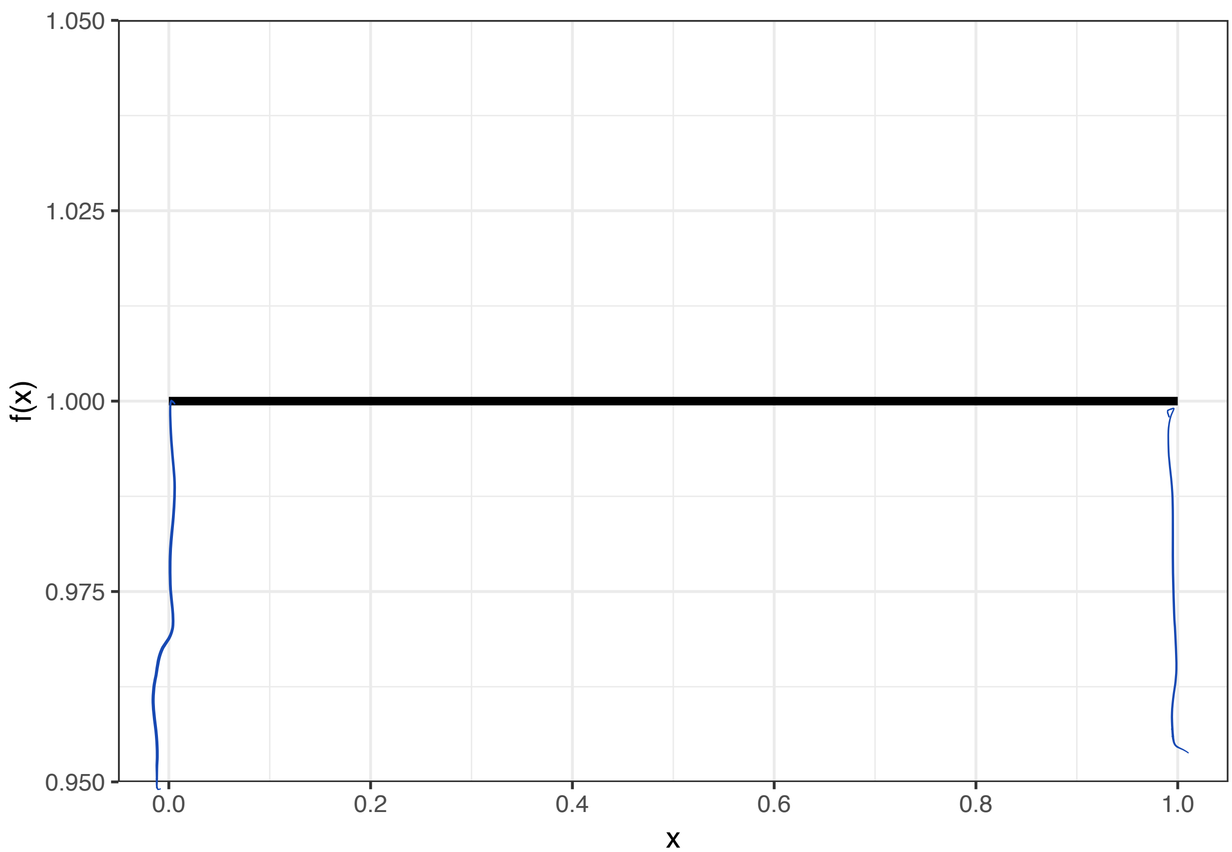
$$\mathcal{R} = [0, 1]$$

$$f(x) = 1 \text{ for } x \in \mathcal{R}$$

$$F(y) = y \text{ for } y \in \mathcal{R}$$

$$\text{E}[X] = 1/2, \ \text{Var}(X) = 1/12$$

Uniform (Continuous) PDF



Uniform (Continuous) in R

```
> str(dunif)
function (x, min = 0, max = 1, log = FALSE)
```

```
> str(punif)
function (q, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
```

```
> str(qunif)
function (p, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
```

```
> str(runif)
function (n, min = 0, max = 1)
```

Exponential

Models a time to failure and has a “memoryless property”.

$$X \sim \text{Exponential}(\lambda)$$

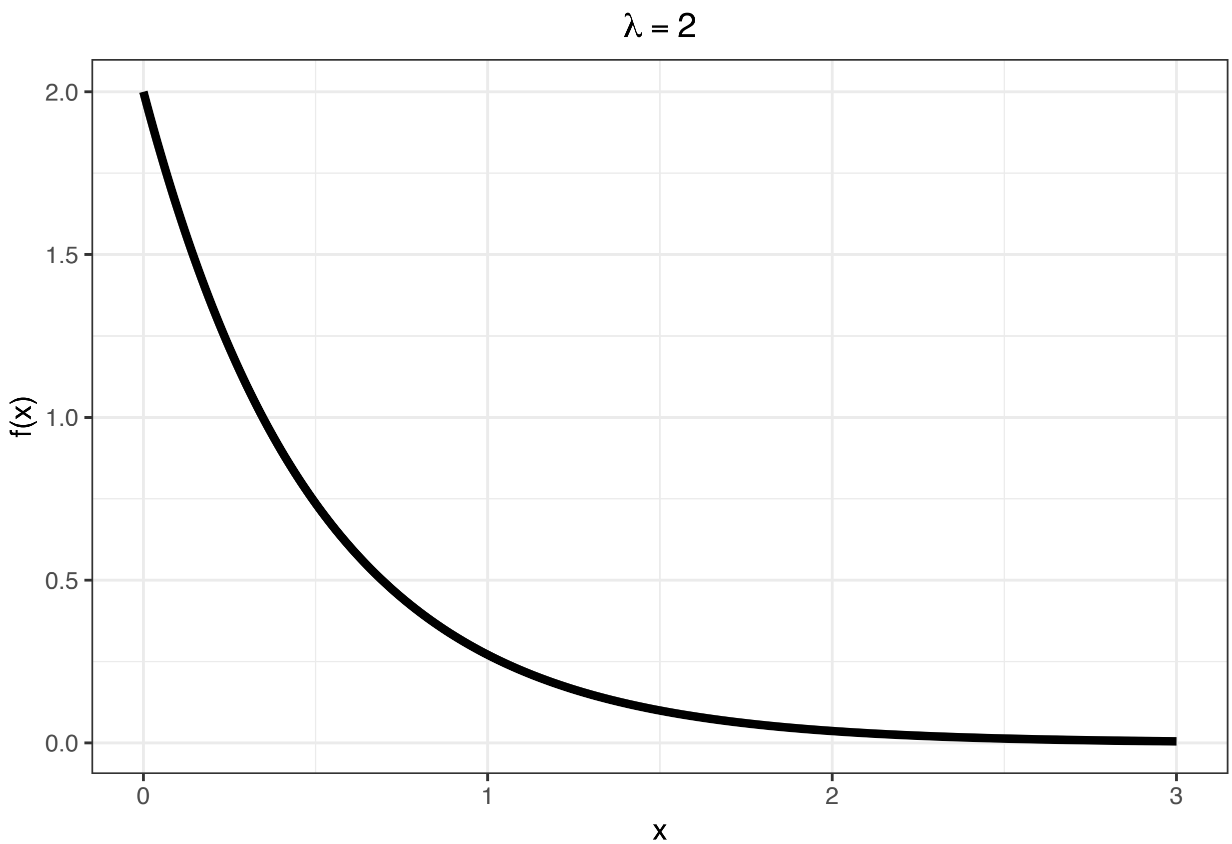
$$\mathcal{R} = [0, \infty)$$

$$f(x; \lambda) = \lambda e^{-\lambda x} \text{ for } x \in \mathcal{R}$$

$$F(y; \lambda) = 1 - e^{-\lambda y} \text{ for } y \in \mathcal{R}$$

$$\text{E}[X] = \frac{1}{\lambda}, \ \text{Var}(X) = \frac{1}{\lambda^2}$$

Exponential PDF



Exponential in R

```
> str(dexp)
function (x, rate = 1, log = FALSE)
```

```
> str(pexp)
function (q, rate = 1, lower.tail = TRUE, log.p = FALSE)
```

```
> str(qexp)
function (p, rate = 1, lower.tail = TRUE, log.p = FALSE)
```

```
> str(rexp)
function (n, rate = 1)
```

Beta

Yields values in $(0, 1)$, so often used to generate random probabilities, such as the p in Bernoulli(p).

$$X \sim \text{Beta}(\alpha, \beta) \text{ where } \alpha, \beta > 0$$

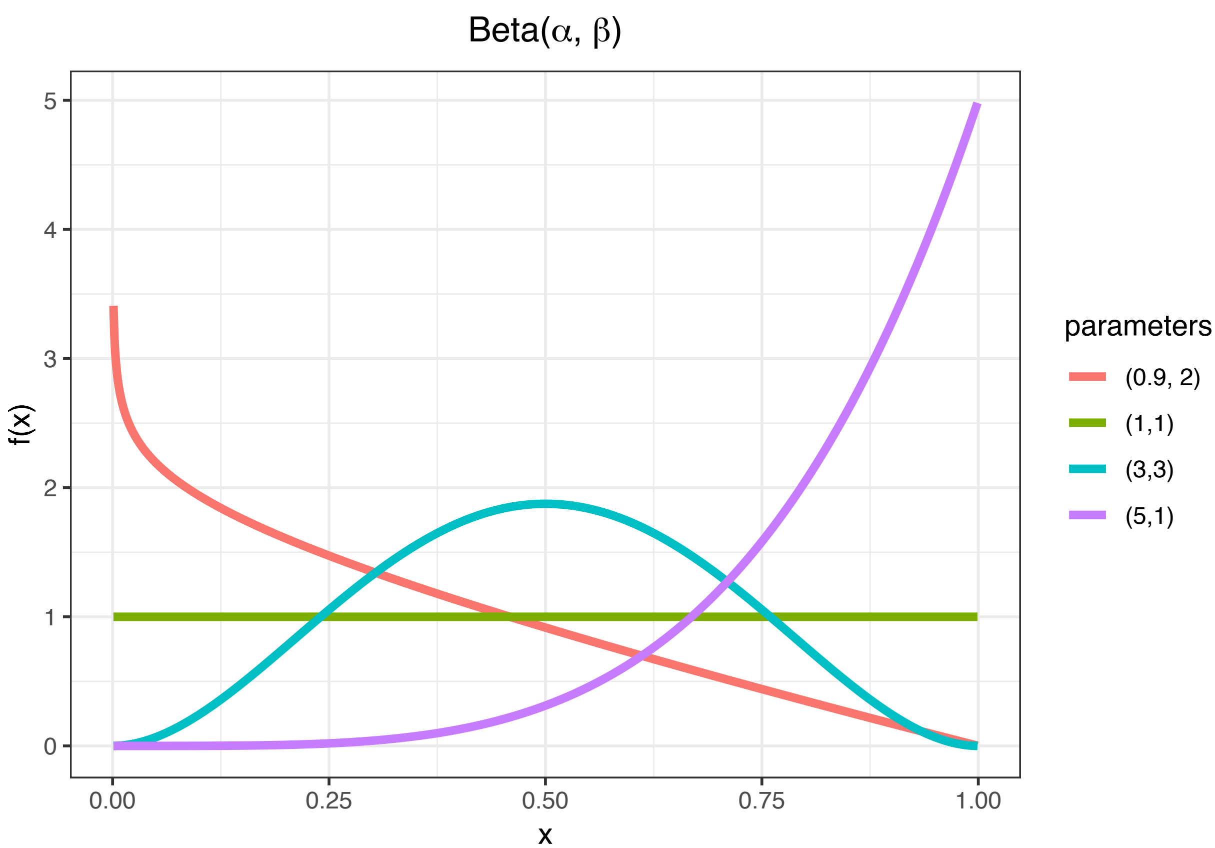
$$\mathcal{R} = (0, 1)$$

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } x \in \mathcal{R}$$

where $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$.

$$\text{E}[X] = \frac{\alpha}{\alpha + \beta}, \ \text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Beta PDF



Beta in R

```
> str(dbeta) #shape1=alpha, shape2=beta
function (x, shape1, shape2, ncp = 0, log = FALSE)
```

```
> str(pbeta)
function (q, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
```

```
> str(qbeta)
function (p, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
```

```
> str(rbeta)
function (n, shape1, shape2, ncp = 0)
```

Normal

Due to the Central Limit Theorem (covered later), this "bell curve" distribution is often observed in properly normalized real data.

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

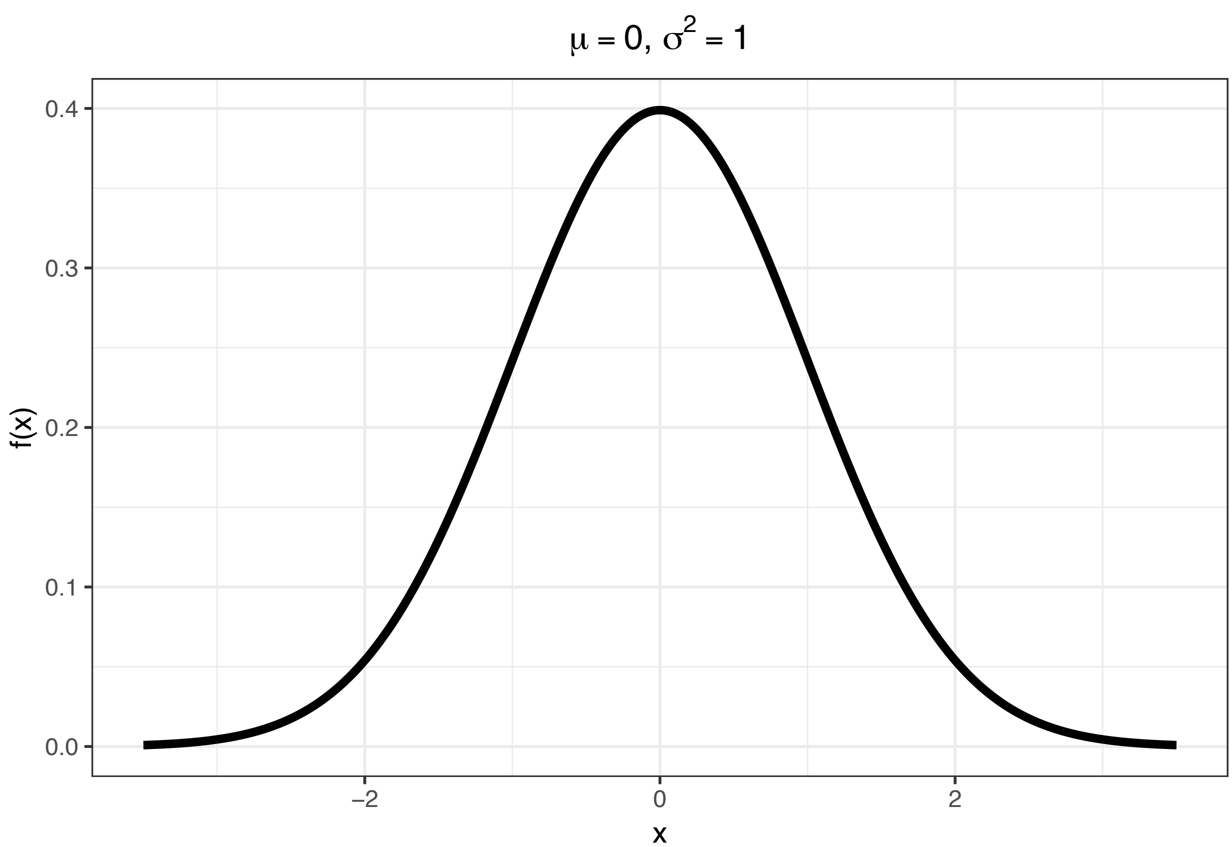
$\sigma = \sqrt{\sigma^2}$

$$\mathcal{R} = (-\infty, \infty)$$

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for } x \in \mathcal{R}$$

$$\text{E}[X] = \mu, \ \text{Var}(X) = \sigma^2$$

Normal PDF



Normal in R

```
> str(dnorm) #notice it requires the STANDARD DEVIATION, not the variance
function (x, mean = 0, sd = 1, log = FALSE)
```

```r
> str(pnorm)
function (q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
```

```r
> str(qnorm)
function (p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
```

```r
> str(rnorm)
function (n, mean = 0, sd = 1)
```

$\mu = 0 \qquad \sigma^2 = 1$

pdf:

$$\int \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx = 1$$

$$1 = \int \frac{1}{2\pi} e^{\frac{-x^2 - y^2}{2}} \, dx \, dy$$

$$= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

1. Sums of rv's
2. Convergence of rv's

Sums of rv's

1. $E[a+bX] = a + bE[X]$

   a, b constant

2. $Var(a+bX) = b^2 Var(X)$

3. If $X_1, X_2, X_3, \ldots, X_n$ are independent random variables

   $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] \quad \text{(indep. not required)}$$

   $$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

4. If $X_1, X_2, \ldots, X_n$ are dependent rv's:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i,j} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

Means of RVs

$X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

μ

$$E[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^{n} E[X_i]$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[X]$$

$$= E[X] = \mu$$

$$Var(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X)$$

$$= \frac{Var(X)}{n} = \frac{\sigma^2}{n}$$

Central Limit Theorem

$X_1, X_2, \ldots, X_n$ are i.i.d. rv's with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$.

Then as $n \to \infty$

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} Normal(0, \sigma^2)$$

$$\frac{\bar{X}_{40} - 6}{\sqrt{6/40}} \quad N(0,1)$$

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Normal(0, 1)$$

$$Pr\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le y\right) \xrightarrow{n\to\infty} P(Z \le y)$$

where $Z \sim Normal(0,1)$

Fact: $Z = \dfrac{X - E[X]}{\sqrt{Var(X)}}$

$$\Rightarrow \quad E[Z] = 0 \quad Var(Z) = 1$$

Define $a = \dfrac{-E[X]}{\sqrt{Var(X)}}$

$b = \dfrac{1}{\sqrt{Var(X)}}$

$$0.95 = \left(-1.96 \leq \frac{\bar{X}_{40} - \lambda}{\sqrt{\lambda/40}} \leq 1.96 \right)$$

CLT Example

Let $X_1, X_2, \ldots, X_{40}$ be iid Poisson(λ) with $\lambda = 6$.

We will form $\sqrt{40}(\overline{X} - 6)$ over 10,000 realizations and compare their distribution to a Normal(0, 6) distribution.

```
> x <- replicate(n=1e4, expr=rpois(n=40, lambda=6),
+                simplify="matrix")
> x_bar <- apply(x, 2, mean)
> clt <- sqrt(40)*(x_bar - 6)
>
> df <- data.frame(clt=clt, x = seq(-18,18,length.out=1e4),
+                  y = dnorm(seq(-18,18,length.out=1e4),
+                            sd=sqrt(6)))
```

```
> ggplot(data=df) +
+   geom_histogram(aes(x=clt, y=..density..), color="blue",
+                  fill="lightgray", binwidth=0.75) +
+   geom_line(aes(x=x, y=y), size=1.5)
```