Examples

The bionomial distribution with parameters (n, p) has sample space $\{0, 1, \ldots, n\}$ and point probabilities

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$$

The uniform distribution on $\{1, \ldots, n\}$ has point probabilities

$$p(k) = \frac{1}{n}, \quad k = 1, \ldots, n.$$

Mean and variance

If P is a probability measure on a discrete set $E \subseteq \mathbb{R}$ with point probabilities $p(x)$ for $x \in E$ we define the mean and the variance as

$$\mu = \sum_{x \in E} x p(x)$$

and

$$\sigma^2 = \sum_{x \in E} (x - \mu)^2 p(x).$$

The former is only meaningful if

$$\sum_{x \in E} |x| p(x) < \infty$$

and the latter only if

$$\sum_{x \in E} x^2 p(x) < \infty.$$

Uniform distribution

The uniform distribution on $\{1, \ldots, n\}$ has mean

$$\mu = \frac{n+1}{2}$$

and variance

$$\sigma^2 = \frac{(n+1)(n-1)}{12}.$$

Distribution functions

If P is a probability measure on $\mathbb{R}$ the Distribution function is defined as

$$F(x) = P((-\infty, x])$$

for $x \in \mathbb{R}$.

How does such a function look? What are the general characteristics of a distribution function?

Characterization

A distribution function $F : \mathbb{R} \to [0,1]$ satisfies the following properties

(i) F is increasing.

(ii) $F(x) \to 0$ for $x \to -\infty$, $F(x) \to 1$ for $x \to \infty$.

(iii) F is right continuous.

Important characterization: Any function $F : \mathbb{R} \to [0,1]$ satisfying the properties (i)-(iii) above is the distribution function for a unique probability measure.

Discrete distributions

A binomial distribution on $\{0, \ldots, 10\}$ and probability parameter $p = 1/2$ has point probabilities, which we can get from R.

```
> pb <- dbinom(c(0:10), 10, 1/2)
```

Compute the mean and variance for this binomial distribution.

We can also simulate 100 Binomial experiments with probability parameter 1/2

```
> tmp <- rbinom(100, 10, 1/2)
```

Use `mean` and `var` to compute the empirical mean and variance for the resulting 100 simulated variables.

Solutions

Compute the mean and variance for this binomial distribution:

```
> pb <- dbinom(c(0:10), 10, 1/2)
> mu <- sum((0:10) * pb)
> mu

[1] 5

> sum((0:10 - mu)^2 * pb)

[1] 2.5


> tmp <- rbinom(100, 10, 1/2)
> mean(tmp)

[1] 5.19

> var(tmp)
```


Densities

If $f : \mathbb{R} \to [0, \infty)$ satisfies that

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

we call f a (probability) density function. The corresponding probability measure is given by

$$P(A) = \int_A f(x)\mathrm{d}x$$

and has distribution function

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

If a distribution function F is differentiable then there is a density

$$f(x) = F'(x).$$

Distribution functions

Plot the graph (use `plot` or `curve`) for the function

```
> F <- function(x) 1 - x^(-0.3) * exp(-0.4 * (x - 1))
```

for $x \in [1, \infty)$. Argue that it is a distribution function.

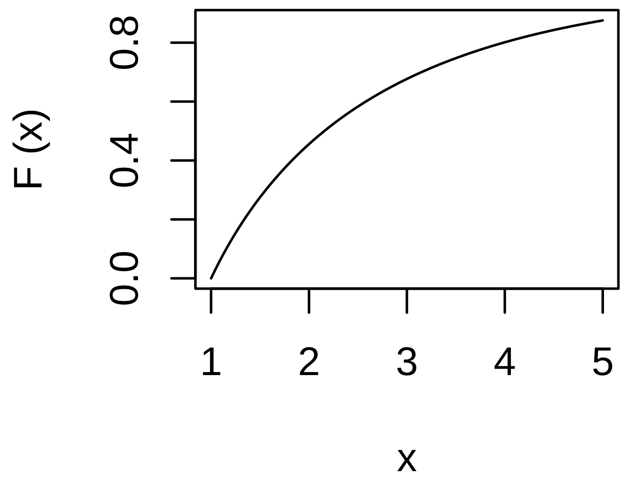
Define

```
> f <- function(x) x^3 * exp(-x)/6
```

for $x \in [0, \infty)$ and use `integrate(f,0,Inf)` to verify that f is a density.

How can you use `integrate` to create the corresponding distribution function?

Solutions




```
> F <- function(x) 1 - x^(-0.3) * exp(-0.4 * (x - 1))
> curve(F, 1, 5)
```


Solutions – continued

```
> f <- function(x) x^3 * exp(-x)/6
> integrate(f, 0, Inf)$value
```

```
[1] 1
```

```
> F.simple <- function(x) integrate(f, 0, x)$value
> F <- function(x) sapply(x, F.simple)
```

The latter `F` works correctly when given a vector input, the former `F.simple` does not.

Mean and variance

If P is a probability measure on $\mathbb{R}$ given by the density f we define the mean

$$\mu = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

and the variance

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \mathrm{d}x.$$

The former is meaningful if

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x < \infty$$

and the latter if

$$\int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x < \infty.$$

Mean and variance

Compute the mean and variance of the distribution with density

```
> f <- function(x) x^3 * exp(-x)/6
```

using `integrate`.

Then compute the mean and variance for the distribution with distribution function

```
> F <- function(x) 1 - x^(-0.3) * exp(-0.4 * (x - 1))
```