Continuous Random Variables
Expected Values and Moments

Statistics 110

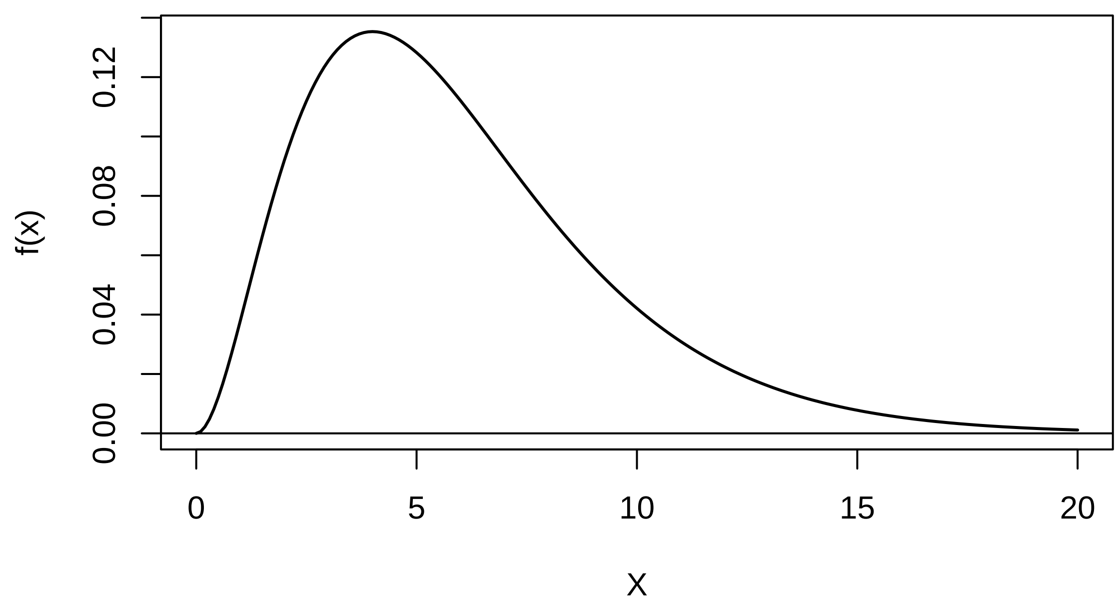
Summer 2006

Continuous Random Variables

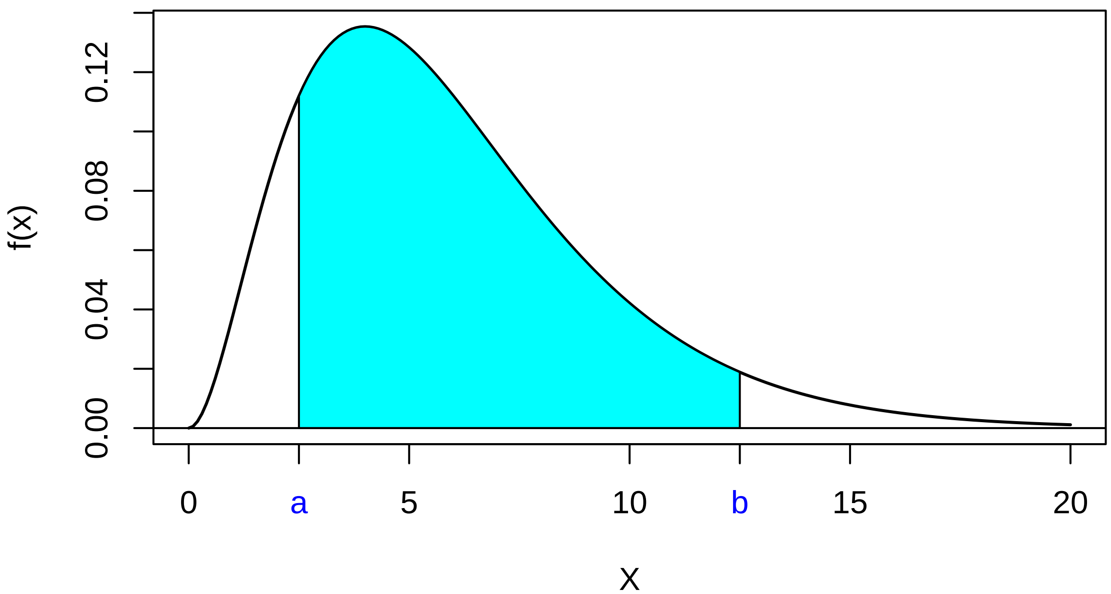
When defining a distribution for a continuous RV, the PMF approach won't quite work since summations only work for a finite or a countably infinite number of items. Instead they are based on the following

Definition: Let X be a continuous RV. The **Probability Density Function** (PDF) is a function $f(x)$ on the range of X that satisfies the following properties:

- $f(x) \geq 0$
- f is piecewise continuous
- $\int_{-\infty}^{\infty} f(x)dx = 1$



For any $a < b$, the probability that $P[a < X < b]$ is the area under the density curve between a and b.



$$P[a < X < b] = \int_a^b f(x)dx$$

Note that $f(a)$ is **NOT** the probability of observing $X = a$ as

$$P[X = a] = \int_a^a f(x)dx = 0$$

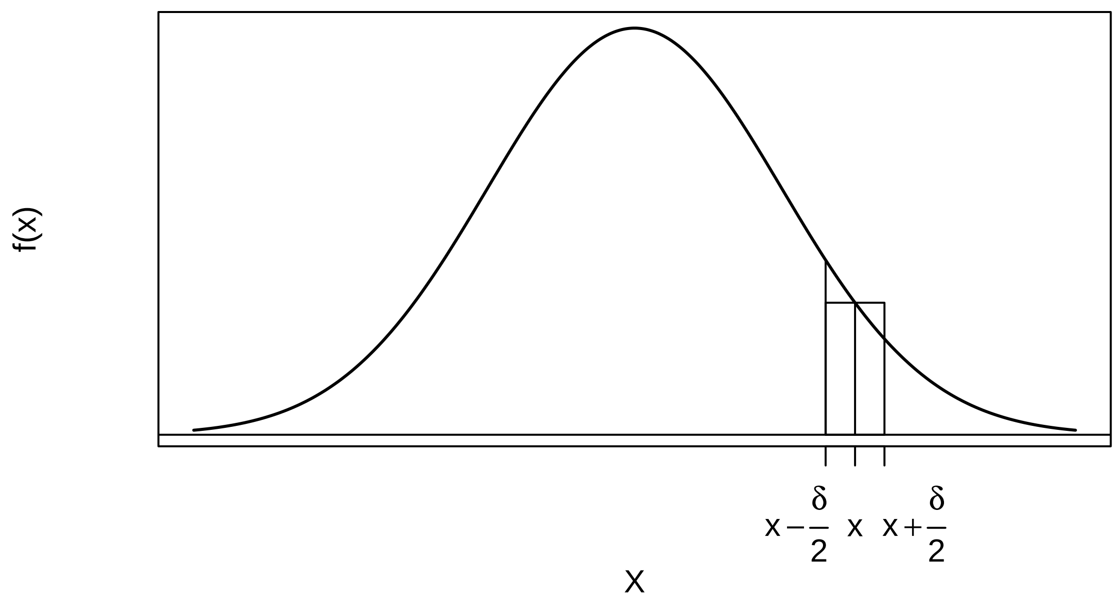
Thus the probability that a continuous RV takes on any particular value is 0. (While this might seem counterintuitive, things do work properly.) A consequence of this is that

$$P[a < X < b] = P[a \leq X < b] = P[a < X \leq b] = P[a \leq X \leq b]$$

for continuous RVs. Note that this won't hold for discrete RVs.

Note that for small δ, if f is continuous at x

$$P\left[x - \frac{\delta}{2} \leq X \leq x + \frac{\delta}{2}\right] = \int_{x-\frac{\delta}{2}}^{x+\frac{\delta}{2}} f(u)du \approx f(x)\delta$$



So the probability of seeing an outcome in a small interval around x is proportional to $f(x)$. So the PDF is giving information of how likely an observation at x is.

As with the PMF and the CDF for discrete RVs, there is a relationship between the PDF, $f(x)$, and the CDF, $F(x)$, for continuous RVs

$$\begin{aligned} F(x) &= P[X \leq x] = \int_{-\infty}^{x} f(u)du \\ f(x) &= F'(x) \end{aligned}$$

assuming that f is continuous at x.

Based on this relationship, the probability for any reasonable event describing a RV can determined with the CDF as the probability of any interval satisfies

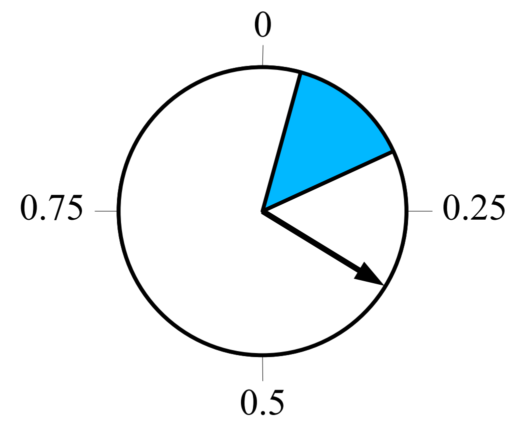
$$P[a < X \leq b] = F(b) - F(a)$$

Note that this is slightly different than the formula given on page 47. The above holds for any RV (discrete, continuous, mixed). The form given on page 47

$$P[a \leq X \leq b] = F(b) - F(a)$$

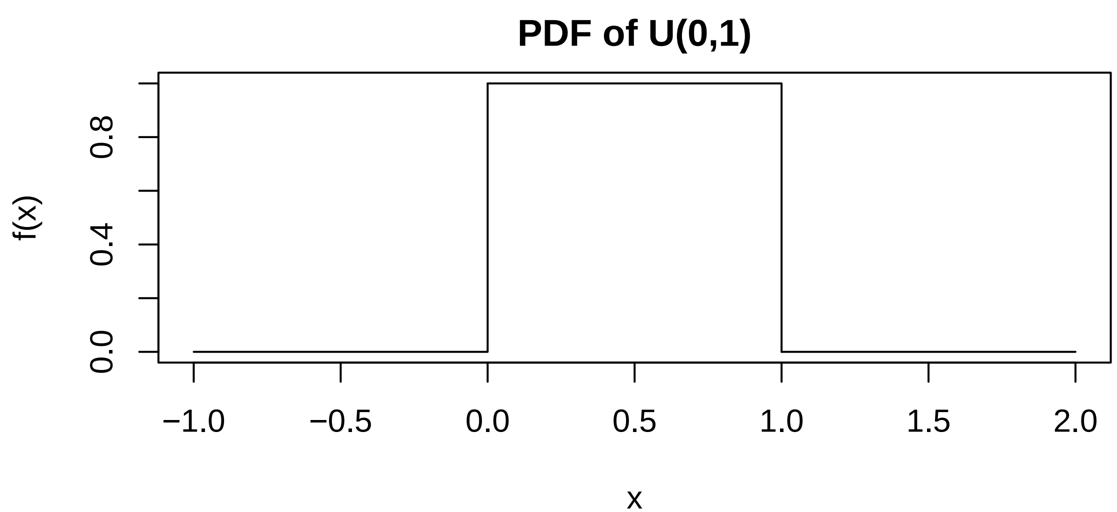
only holds for continuous RVs.

Example: Uniform RV on [0,1] (Denoted $X \sim U(0,1)$)



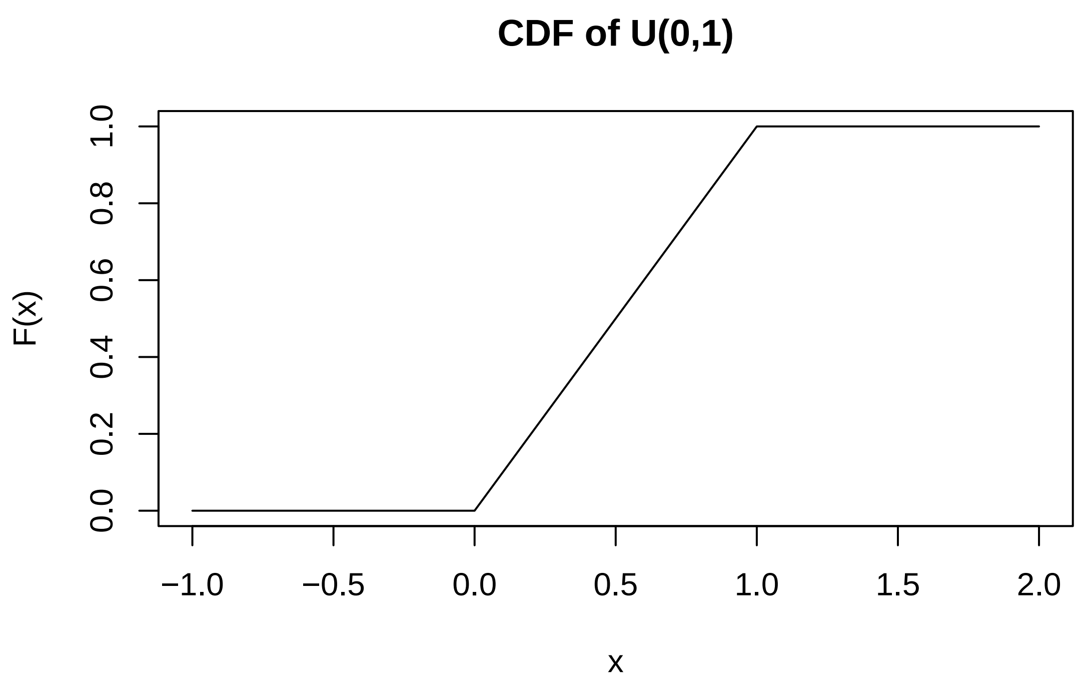
What most people think of when we say pick a number between 0 and 1. Any real number in the interval is possible and equally likely, implying that any interval of length h must have the same probability (which needs to be h). The PDF for X then must be

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$



The CDF for a $U(0,1)$ is

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$



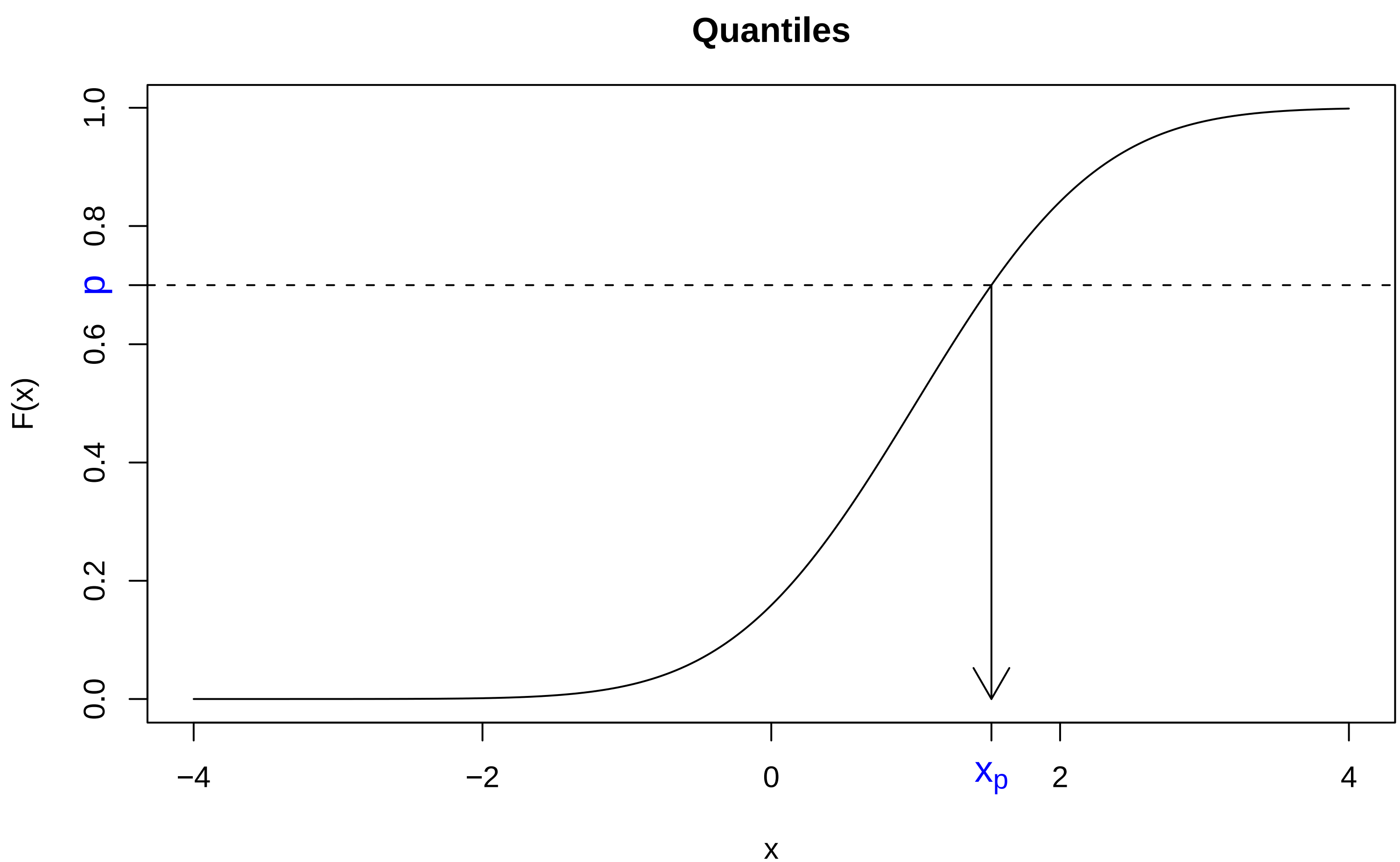
One way to think of the CDF is that you give a value of the RV and it gives a probability associated with it (i.e. $P[X \leq x]$). It can also be useful to go the other way. Give a probability and figure out which value of the RV is associated with it.

Lets assume that F is continuous and strictly increasing in some interval I (i.e. $F = 0$ to the left of I and $F = 1$ to the right of I) (note I might be unbounded). Under these assumptions the inverse function F^{-1} is well defined ($x = F^{-1}(y)$ if $F(x) = y$).

Definition: The pth **Quantile** of the distribution F is defined to be the value x_p such that

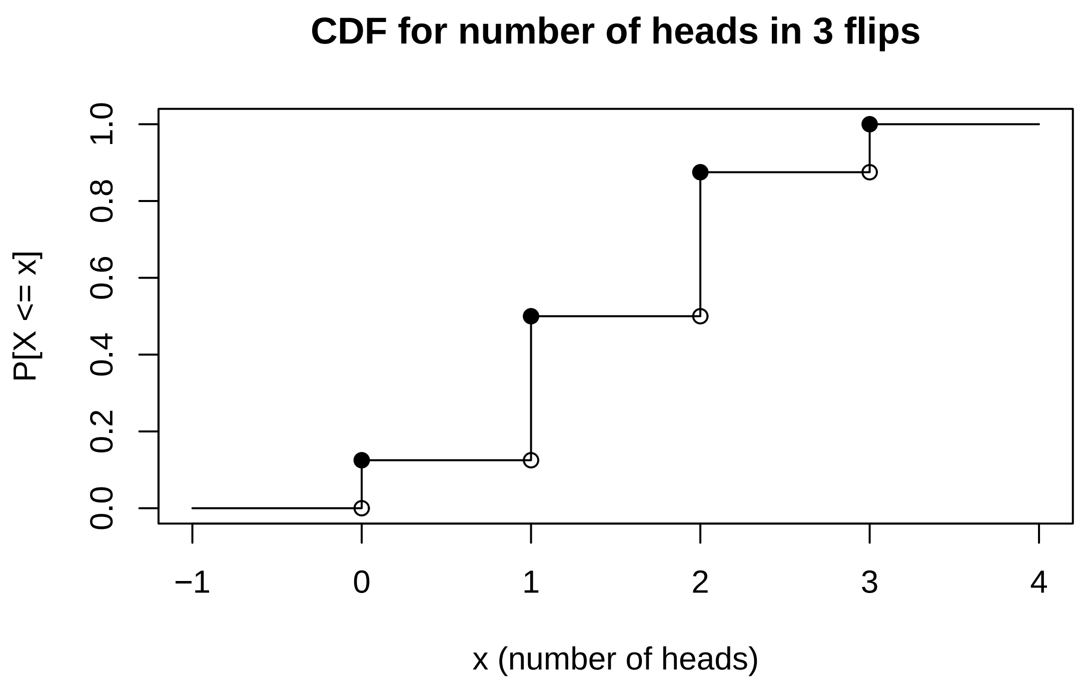
$$F(x_p) = p \quad \text{or} \quad P[X \leq x_p] = p$$

Under the above assumptions $x_p = F^{-1}(p)$.



Special cases of interest of the **Median** ($p = \frac{1}{2}$) and the lower and upper **Quartiles** ($p = \frac{1}{4}$ and $= \frac{3}{4}$)

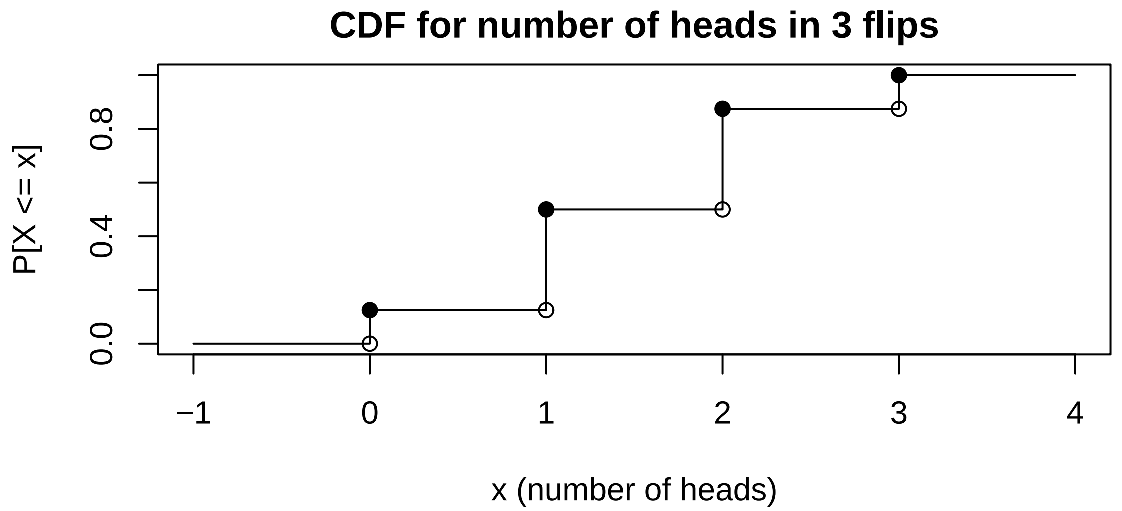
Note: Defining quantiles for discrete distributions is a bit tougher since the CDF doesn't take all values between 0 and 1 (due to the jumps)

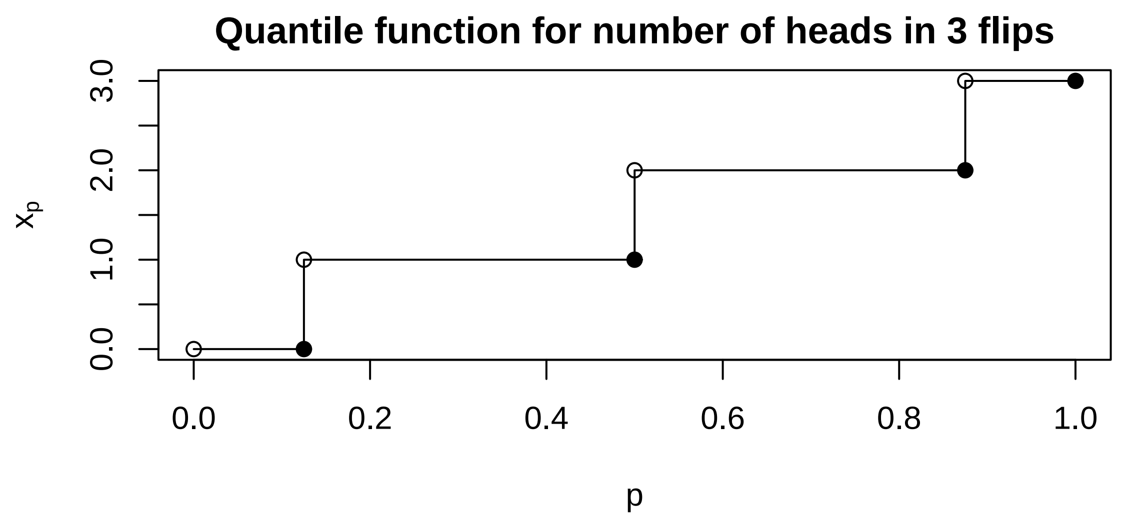


The definition above can be extended to solving the simultaneous equations

$$P[X \leq x_p] \geq p \quad \text{and} \quad P[X < x_p] \leq p$$

This can be though of as the place where the CDF jumps from below p to above p





Expected Values and Moments

Definition: The **Expected Value** of a continuous RV X (with PDF $f(x)$) is

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

assuming that $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$.

The expected value of a distribution is often referred to as the mean of the distribution.

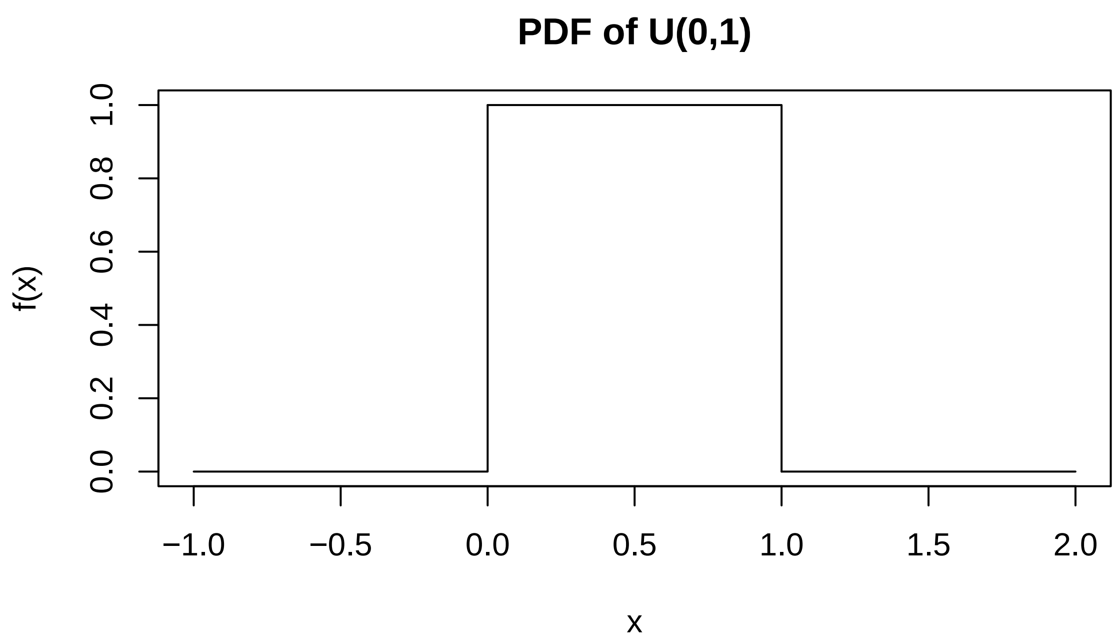
As with the discrete case, the absolute integrability is a technical point, which if ignored, can lead to paradoxes.

For an example of a continuous RV with infinite mean, see the Cauchy distribution (Example G, page 114)

As with the discrete case, $E[X]$ can be thought as a measure of center of the random variable.

For example, when $X \sim U(0,1)$

$$E[X] = \int_0^1 x dx = 0.5$$



Not surprisingly, expectations of functions of continuous RVs satisfy the expected relationship

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

For example, if $X \sim U(0, 1)$,

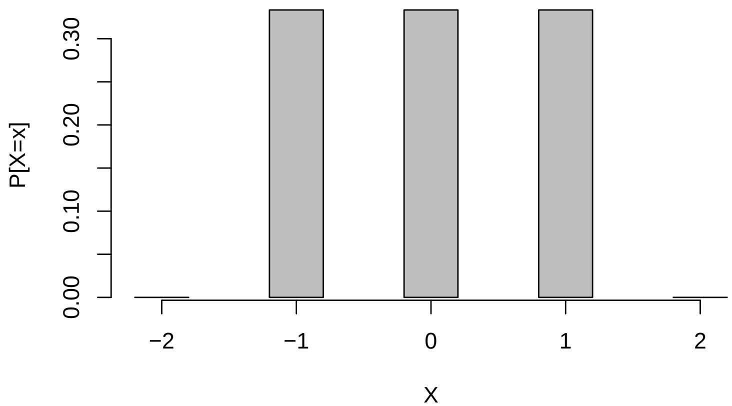
$$E[X^2] = \int_0^1 x^2 dx = \frac{1}{3}$$

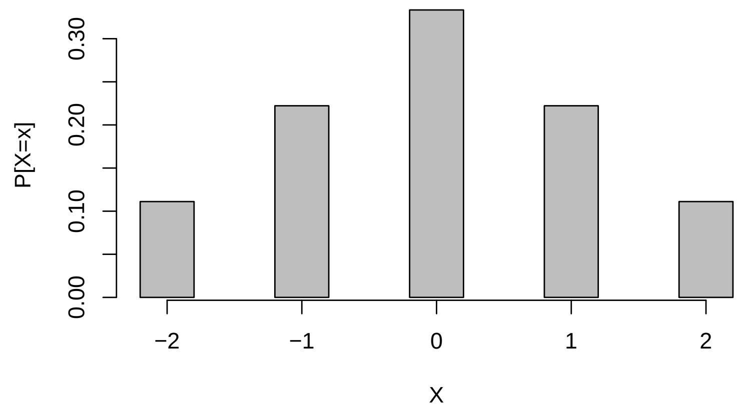
This is often easier than figuring out the PDF of $Y = g(X)$ and applying the definition as there is often some work to figure out the PDF of Y. (Which we will do later, it does have its uses)

As with discrete RVs, $g(E[X]) \neq E[g(X)]$ in most cases. However, with a linear transformation $Y = a + bX$

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

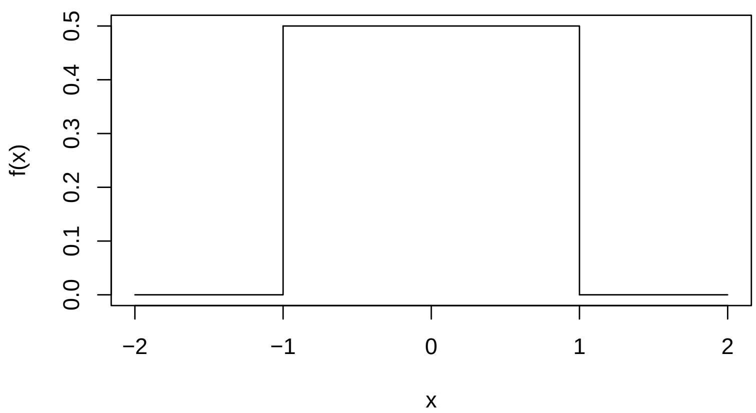
Spread of a RV

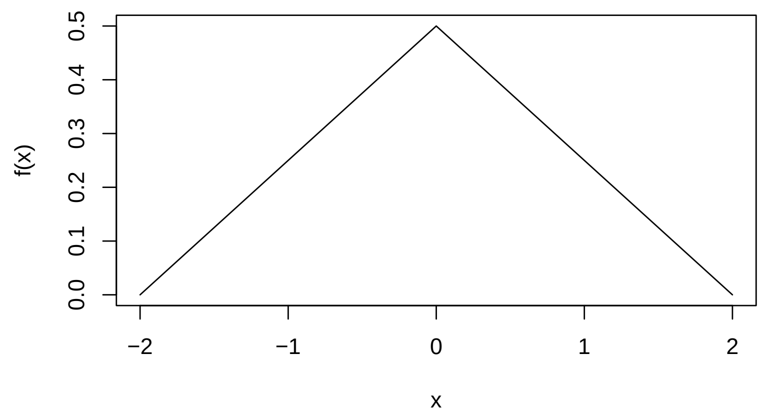




x	-1	0	1
$p(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

x	-2	-1	0	1	2
$p(x)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$





$$f(x) = \begin{cases} 0.5 & -1 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases} \qquad f(x) = \begin{cases} 0.5 + \frac{x}{4} & -2 \leq x \leq 0 \\ 0.5 - \frac{x}{4} & 0 \leq x \leq 2 \\ 0 & \text{Otherwise} \end{cases}$$

All these distributions have $E[X] = 0$ but the right hand side in each case has a bigger spread. A common measure of spread is the Standard Deviation

Definition: Let $\mu = E[X]$, then the **Variance** of the random variable X is

$$\mathrm{Var}(X) = E[(X - \mu)^2]$$

provided the expectation exists.

The **Standard Deviation** of X is

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

For a discrete RV,

$$\mathrm{Var}(X) = \sum_i (x_i - \mu)^2 p(x_i)$$

For a continuous RV

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

The variance measures the expected squared difference of an observation from the mean. While the interpretation of the standard deviation isn't quite easy, it can be thought of a measure of the typical spread of a RV.

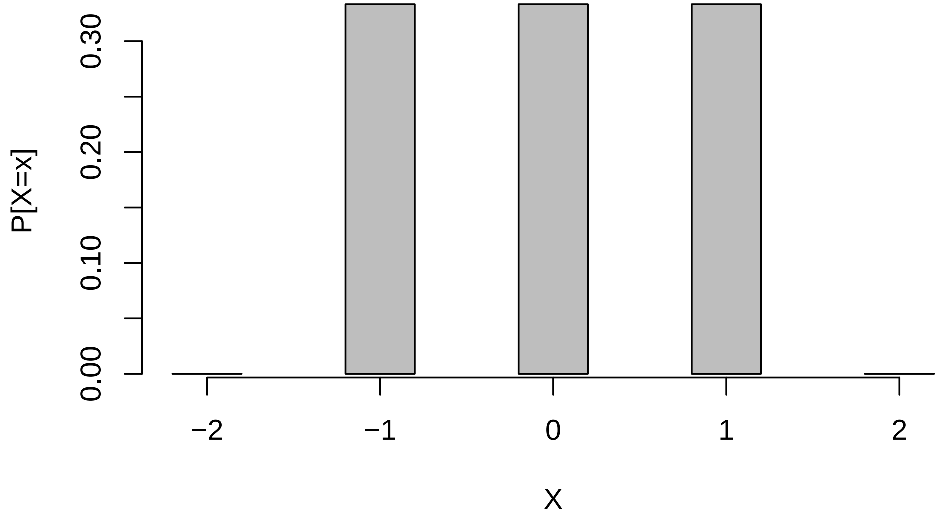
It can be shown that, assuming that the variance exists,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

This form is often useful for calculation purposes.

Notation: The variance is often denoted by σ^2 and the standard deviation by σ.

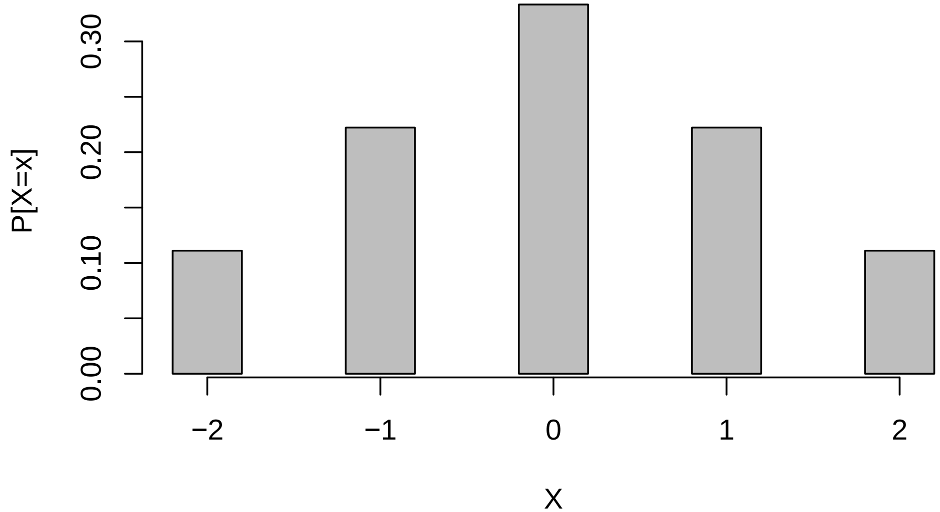
For the examples



x	-1	0	1
$p(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$\mathrm{Var}(X) = (-1-0)^2\frac{1}{3} + (0-0)^2\frac{1}{3} + (1-0)^2\frac{1}{3} = \frac{2}{3}$$

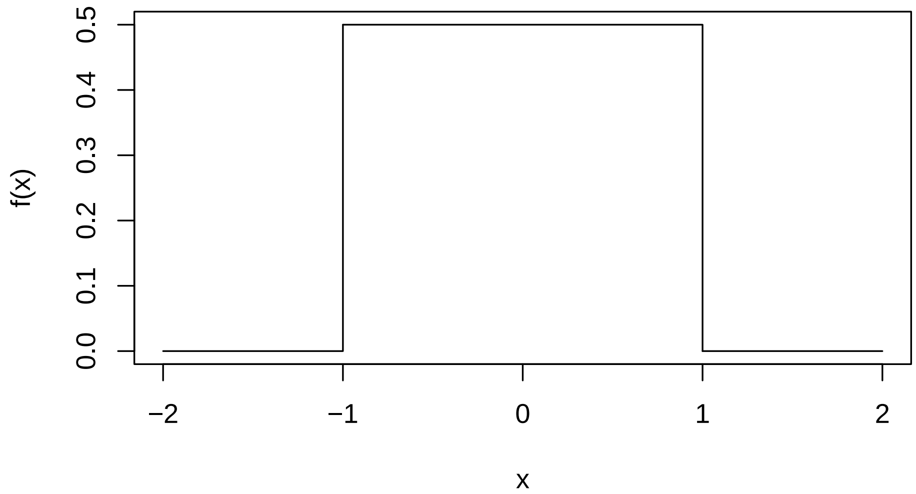
$$\mathrm{SD}(X) = \sqrt{\frac{2}{3}} = 0.8165$$



x	-2	-1	0	1	2
$p(x)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

$$
\begin{aligned}
\mathrm{Var}(X) &= (-2-0)^2\frac{1}{9} + (-1-0)^2\frac{2}{9} + (0-0)^2\frac{3}{9} + (1-0)^2\frac{2}{9} + (2-0)^2\frac{1}{9} \\
&= \frac{10}{9}
\end{aligned}
$$

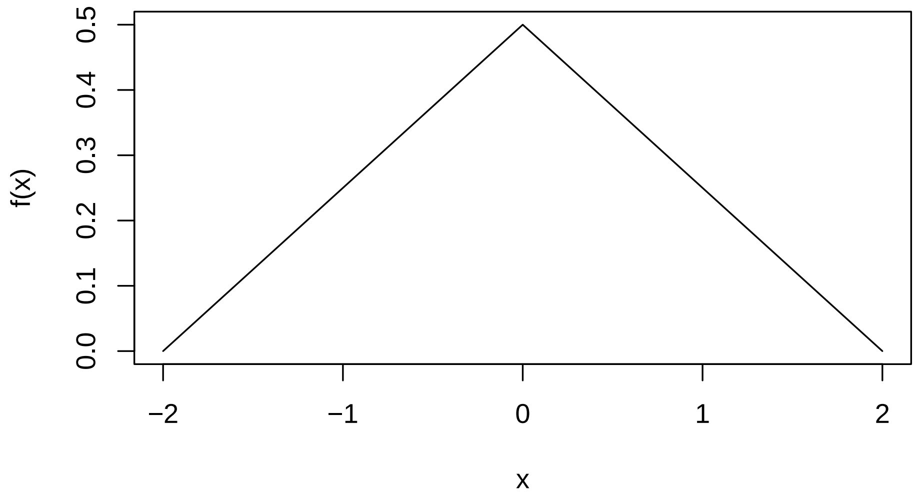
$$\mathrm{SD}(X) = \sqrt{\frac{10}{9}} = 1.0541$$



$$f(x) = \begin{cases} 0.5 & -1 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

$$\mathrm{Var}(X) = \int_{-1}^{1} (x-0)^2 \frac{1}{2} dx = \frac{1}{3}$$

$$\mathrm{SD}(X) = \sqrt{\frac{1}{3}} = 0.5774$$



$$f(x) = \begin{cases} 0.5 + \frac{x}{4} & -2 \leq x \leq 0 \\ 0.5 - \frac{x}{4} & 0 \leq x \leq 2 \\ 0 & \text{Otherwise} \end{cases}$$

$$\mathrm{Var}(X) = 2\int_0^2 (x-0)^2 (0.5 - \frac{x}{4}) dx = \frac{4}{3}$$

$$\mathrm{SD}(X) = \sqrt{\frac{4}{3}} = 1.1547$$

What is the effect of a linear transformation $(Y = a + bX)$ on the variance and standard deviation?

$$\text{Var}(a + bX) = b^2\text{Var}(X) \qquad \text{SD}(a + bX) = |b|\text{SD}(X)$$

These two results are to be expected. For example, if two possible X values differ by $d = |x_1 - x_2|$, the corresponding Y values differ by $|b|d$, suggesting that we want the standard deviation to scale by a factor of $|b|$. Since the variance measures squared spread, it needs to scale by a factor of b^2.

The factor a not having an effect also makes sense. Adding a to a random variable shifts the location of its distribution, but doesn't changes the distance between corresponding pairs of points.