Continuous RV

- Recall the definition of pmf for a discrete rv. $P(X=x)$. Can we extend this definition to continuous rv’s?

- Uniform random variable: X is equally likely to be any number on [0,1], what is the probability $P(X=0.5)$?

- The probability model for a continuous random variable assigns probabilities to intervals of outcomes rather than to individual outcomes.

- The probability model of X is often described by a smooth curve, which is the probability density function (pdf) of X.

The CDF

- Although the idea of pmd does not extend to the continuous rv's, the idea of cdf still works.

- The cumulative distribution function (cdf) $F(x)$ for a continuous rv X is defined for every number x by

$$F(x) = \mathrm{P}(\mathrm{X} \leq x) = \int_{-\infty}^{x} f(y)dy$$

- $F(x)$ is in fact the probability that a rv X is smaller than x. $F(x)$ increases smoothly as x increases. $F(-\infty) = 0$, $F(+\infty) = 1$.

- It is easy to compute probabilities using $F(x)$.
  - $\mathrm{P}(\mathrm{X} > a) = 1 - F(a)$
  - $\mathrm{P}(a \leq \mathrm{X} \leq b) = F(b) - F(a)$

pdf from cdf

- If X is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every x at which the derivative $F'(x)$ exists, $F'(x) = f(x)$. $f(x)$ is often a smooth curve, which is the probability density function (pdf) of X.

- Let p be a number between 0 and 1. The (100p)th percentile (quantile) of the distribution of a continuous rv X, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)dy$$

- The median of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies .5 = F($\tilde{\mu}$). That is, half the area under the density curve is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$.

PDF

- The probability density function (pdf) of a continuous rv X is a function *f*(*x*) such that for any two numbers *a* and *b* with *a* ≤ *b*,

  $$\mathrm{P}(a \leq \mathrm{X} \leq b) = \int_a^b f(x)dx$$

  The graph of *f*(*x*) is often referred to as the *density curve*.

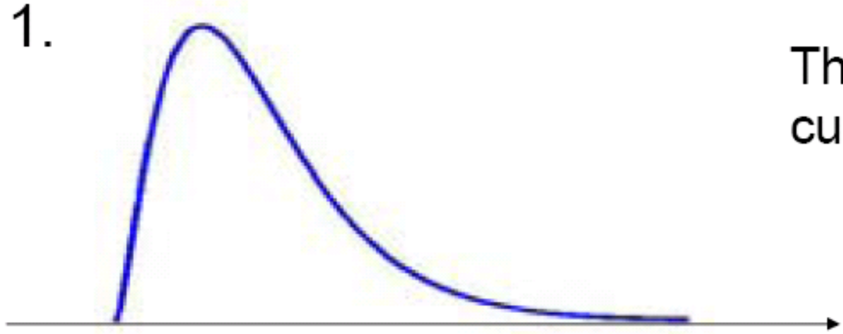
- This means the area under the density curve represents probability!

- Note that 0 ≤ *f*(*x*) for all *x*.

- *f*(*x*)d*x* can be treated as P(X=x)!

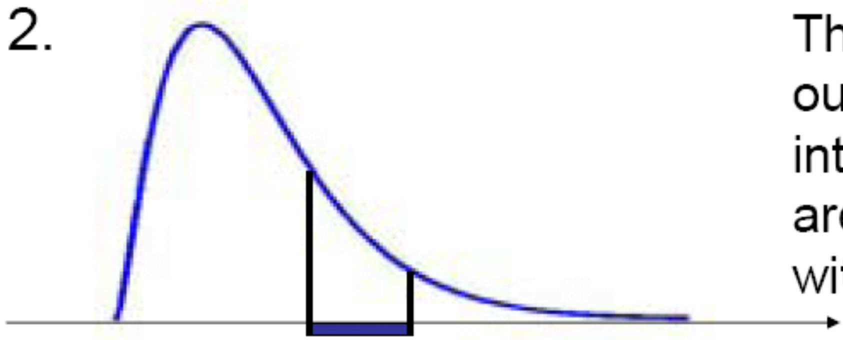
Properties of PDF

1.



The total area under the curve must equal 1.

2.



The probability that the outcome lies in a specific interval is given by the area under the curve within that interval.

Uniform Distribution

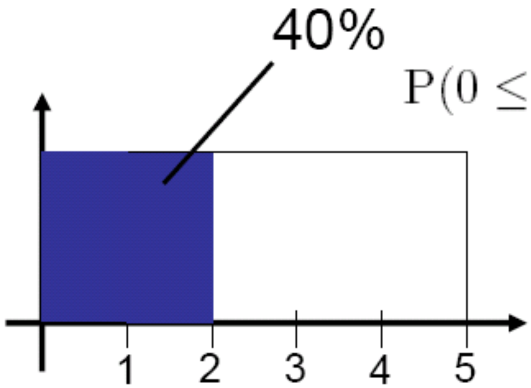
- A continuous rv X is said to have a uniform distribution on the interval [*A*, *B*] if the pdf of X is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

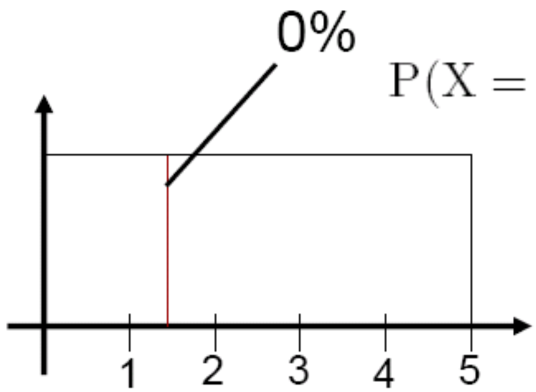
- Verify that this is a proper pdf.
    1. $f(x) \geq 0$ for all x.
    2. Area under $f(x)$ should be equal to 1.

Example

Ex. Suppose a bus arrives equally likely at any time between 7:00 – 7:05 AM. What is the probability it arrives sometime between 7:00 – 7:02 AM?



$$P(0 \leq X \leq 2) = \int_0^2 \frac{1}{5} dx = \frac{2}{5}$$



$$P(X = c) = \lim_{\epsilon \to 0} P(c - \epsilon \leq X \leq c + \epsilon) = \lim_{\epsilon \to 0} \int_{c-\epsilon}^{c+\epsilon} \frac{1}{B - A} dx = 0$$

Expected Values

- Notice that the pdf *f*(*x*) of a continuous distribution is actually playing the role of pmf *p*(*x*) of a discrete distribution.

- Recall that the expected value of a discrete distribution is calculated by

$$\mu_X = \mathrm{E}(\mathrm{X}) = \sum_{x \in D} x \cdot p(x)$$

- Therefore, similarly we can define the expected value of a continuous distribution by

$$\mu_X = \mathrm{E}(\mathrm{X}) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- Take advantage of the *symmetry* of particular distributions, when calculating expectations.

Variance

- With a similar argument as in the discrete case, we can also define the expectation of a function of a continuous rv as well as the variance of a continuous rv.

- Proposition: if X is a continuous rv with pdf *f*(*x*) and *h*(X) is any function of X, then

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

- As a special case of the above proposition, the variance of X is defined by

$$\sigma_X^2 = \mathrm{Var}(\mathrm{X}) = \mathrm{E}(\mathrm{X} - \mathrm{E}(X))^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x)dx$$

The standard deviation (SD) of X is $\sigma_X = \sqrt{\mathrm{Var}(\mathrm{X})}$.

Examples

Ex. Prove for continuous rv X, as in the discrete case, that $Var(X) = E(X^2) - [E(X)]^2$.

Ex. If a stick of length 1 is broken at random into two pieces. What is the expected length of the longer piece?

Properties

- Some properties of mean and variance hold in the continuous case in a similar way as in the discrete case.

- For example, under linear transformation of X, we have

1. $E(aX+b) = aE(X) + b$
2. $Var(aX+b) = a^2 Var(X)$

- Exercise: prove the above formulas rigorously!

Uniform RV

- We call a uniform rv U a standard uniform, if and only if U ~ uniform on [0,1]

- For a standard uniform rv U, we can easily calculate,

$$\mathrm{E}(\mathrm{U}) = \int_0^1 x \cdot 1 dx = \frac{1}{2}$$

$$\mathrm{E}(\mathrm{U}^2) = \int_0^1 x^2 \cdot 1 dx = \frac{1}{3}$$

$$\mathrm{Var}(\mathrm{U}) = \mathrm{E}(\mathrm{U}^2) - [\mathrm{E}(\mathrm{U})]^2 = \frac{1}{12}$$

General Uniform

- Note that a general case of uniform distribution X on [A, B] can be treated as a linear transform of a standard uniform, i.e., $X = (B - A)U + A$.

- Proposition:

> If X is a continuous uniform rv on [A, B], then
> $E(X) = (B + A)/2$, $Var(X) = (B - A)^2/12$

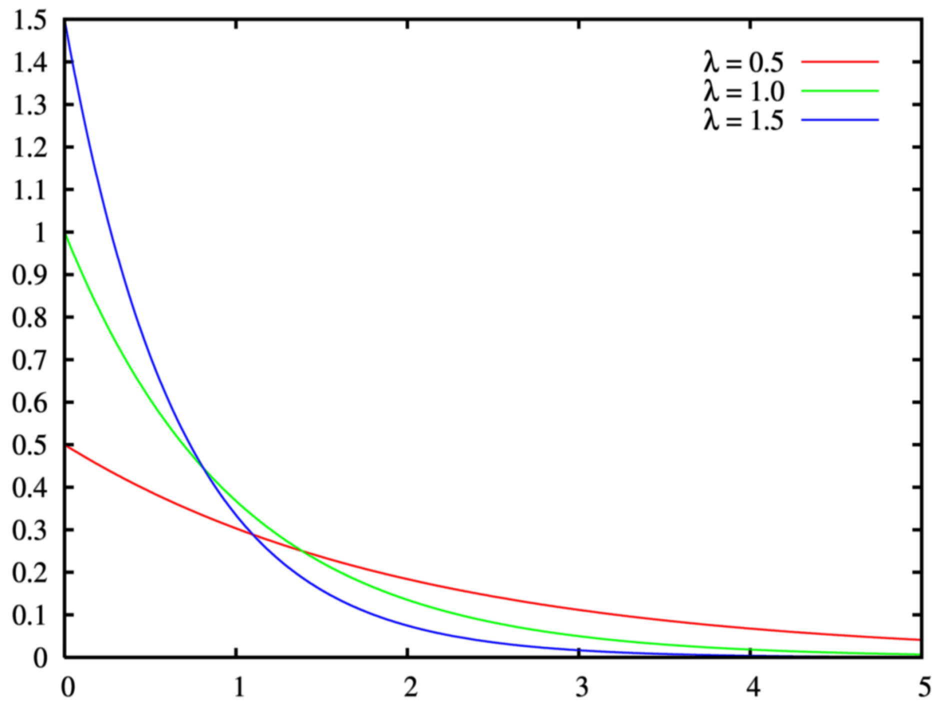
- R command:

```
dunif(x, min=0, max=1),
punif(q, min=0, max=1),
qunif(p, min=0, max=1).
```

The Exponential Distribution

- X is said to have an exponential distribution with parameter λ (λ>0) if the pdf of X is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



More on Exponential

- Note that an exponential rv X can only take positive values. And the cdf of X is

$$F(x;\lambda) = \begin{cases} \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Thus $P(X>x) = 1 - F(x;\lambda) = e^{-\lambda x}$

- Proposition: (proof?)

> If X is an exponential rv with parameter λ, then
> $E(X) = 1/\lambda$, $Var(X) = 1/\lambda^2$

- R command: `dexp(x, lamda=1)`,
  `pexp(q, lamda=1)`,
  `qexp(p, lamda=1)`.