Probability

You perform some experiment (for example roll some dice) whose outcome is not predetermined. Let's say that, in this experiment, you measure some quantity that you call X and that X can take the values $x_1, \cdots, x_n$. If you repeat the experiment you do not always get the same result. But you observe that, after numerous repetitions of the experiment, outcome x_i is observed the fraction p_i of the times. Then we call X a random variable and we say that X takes the value x_i with probability p_i and we write

$$\Pr(X = x_i) = p_i$$

Definition 1.

If a random variable X can only take values in $\{x_k\}$ with the index k running either over a finite set or over the natural numbers, then X is said to be a discrete random variable.

We will later also consider continuous random variables. They take values running over a continuous set, like for example an interval $a \leq x \leq b$. The full definition of a continuous random variable will be given in Definition 11.

Example 2

Suppose that we have a die (that's the singular of "dice") with six sides labelled 1, 2, $\cdots$, 6. The die is "fair", meaning that, when we throw it, each of the six sides is equally likely to end up as the upper face. If we denote by X the number on the upper face, then

$$\Pr(X = 1) = \Pr(X = 2) = \Pr(X = 3) = \Pr(X = 4) = \Pr(X = 5) = \Pr(X = 6) = \frac{1}{6}$$

Example 2

Notation 3.

We use the notation

- $\Pr(a \leq X \leq b)$ to denote the probability that the random variable X takes a value between a and b, and
- $\Pr(X < b)$ to denote the probability that X takes a value strictly less than b, and
- $\Pr(X \neq x)$ to denote the probability that X takes a value other than x, and
- $\Pr(X = a \text{ or } X = b)$ to denote the probability that X takes either the value a or the value b

and so on.

1

Definition 4.

Let X be a discrete random variable.

(a) The set of numbers $\Pr(X = x)$, with x running over all of the values that X takes, is called the *probability distribution* of the random variable X.

(b) If a random variable X can take values x_k with $1 \le k \le n$ and if $\Pr(X = x_k) = \frac{1}{n}$ for all $1 \le k \le n$, then X is said to have a *uniform probability distribution* and is said to be a *uniform discrete random variable*. We write $X = \mathrm{Uniform}(n)$, where n is number of different values that X can take.

(c) If a random variable X can only take two values x_1 and x_2 (for example if X is the result of a coin toss with x_1 representing heads and x_2 representing tails) and if $\Pr(X = x_1) = \Pr(X = x_2) = \frac{1}{2}$, then X is called a *Bernoulli random variable* with parameter $^1/_2$. We write $X = \mathrm{Ber}(^1/_2)$.

Example 5

Now suppose that we have two fair dice, one red and one blue, each with six sides labelled $1, 2, \cdots, 6$. Suppose further that the two dice are *independent*, meaning that the result of throwing one of the dice has no effect on the result of throwing the other one. Denote by X_r the random variable giving the number showing on the upper face of the red die and by X_b the random variable giving the number showing on the upper face of the blue die. Then, there are six equally likely possible outcomes for X_r, namely

$$\{ \, X_r = j \mid j = 1, 2, 3, 4, 5, 6 \, \}$$

Similarly, there are six equally likely possible outcomes for X_b, namely

$$\{ \, X_b = k \mid k = 1, 2, 3, 4, 5, 6 \, \}$$

If we now throw both dice, there are 36 equally likely possible outcomes, namely

$$\{ \, X_r = j, \text{ and } X_b = k \mid j = 1, 2, 3, 4, 5, 6, \ k = 1, 2, 3, 4, 5, 6 \, \} \tag{1}$$

Because each of these outcomes are equally likely,

$$\Pr\left(X_r = j \text{ and } X_b = k \right) = \frac{1}{36}$$

for all $1 \le j \le 6$ and $1 \le k \le 6$.

Now suppose that we are interested in the sum of the two numbers appearing on the upper faces of the two dice. That is, we are interested in the random variable $X = X_r + X_b$. This random variable takes integer values between 2 (since $X_r \ge 1$ and $X_b \ge 1$) and 12 (since $X_r \le 6$ and $X_b \le 6$).

March 7, 2015

- Now $X = 2$ if and only if $X_r = X_b = 1$ and so $\Pr(X = 2) = \frac{1}{36}$.

- For X to take the value 3, we need either $X_r = 1$, $X_b = 2$ or $X_r = 2$, $X_b = 1$. That is, $X = 3$ is given by two of the thirty six equally likely outcomes in (1) and so $\Pr(X = 3) = \frac{2}{36} = \frac{1}{18}$.

- For X to take the value 4, we need (X_r, X_b) to be one of $(1, 3)$, $(2, 2)$, $(3, 1)$. So $X = 4$ is given by three of the thirty six equally likely outcomes in (1) and so $\Pr(X = 4) = \frac{3}{36} = \frac{1}{12}$.

and so on. Here is the full probability distribution for X.

$$\Pr(X = 2) = \frac{1}{36} \qquad \Pr(X = 3) = \frac{2}{36} \qquad \Pr(X = 4) = \frac{3}{36}$$

$$\Pr(X = 5) = \frac{4}{36} \qquad \Pr(X = 6) = \frac{5}{36} \qquad \Pr(X = 7) = \frac{6}{36}$$

$$\Pr(X = 8) = \frac{5}{36} \qquad \Pr(X = 9) = \frac{4}{36} \qquad \Pr(X = 10) = \frac{3}{36}$$

$$\Pr(X = 11) = \frac{2}{36} \qquad \Pr(X = 12) = \frac{1}{36}$$

As a check, note that

$$\sum_{k=2}^{12} \Pr(X = k) = 1$$

which is exactly what we expect — X must take one of the values between 2 and 12.

Example 5

Here are the basic properties that all discrete random variables obey.

Theorem 6.

Let X be a discrete random variable that takes values x_k with k running over some index set $\mathcal{I}$. Then

(a) $0 \le \Pr(x_k) \le 1$ for each $k \in \mathcal{I}$.

(b) $\sum_{k \in \mathcal{I}} \Pr(X = x_k) = 1$. That is X must take the value x_k for some $k \in \mathcal{I}$.

(c) If $j \in \mathcal{I}$, then $\Pr(X = x_j) = 1 - \Pr(X \ne x_j) = 1 - \sum_{\substack{k \in \mathcal{I} \\ k \ne j}} \Pr(X = x_k)$.

(d) $\Pr(X < x) = 1 - \Pr(X \ge x)$ and $\Pr(X \le x) = 1 - \Pr(X > x)$

(e) If $a \ne b$, then $\Pr(X = a \text{ or } X = b) = \Pr(X = a) + \Pr(X = b)$.

 March 7, 2015

> **Definition 7.**
>
> The *cumulative distribution function* (often shortened to CDF) of a random variable X is the function
> $$F(x) = \Pr(X \le x)$$
> In particular, if X is a discrete random variable that takes values x_k with k running over some index set $\mathcal{I}$, then its cumulative distribution function is
> $$F(x) = \sum_{\substack{k \in \mathcal{I} \\ x_k \le x}} \Pr(X = x_k)$$
> The notation " $\displaystyle\sum_{\substack{k \in \mathcal{I} \\ x_k \le x}}$ " indicates that we sum over all values of the index k that are in the index set $\mathcal{I}$ and that obey $x_k \le x$.

Example 8 (Example 2, continued.)

Suppose that, as in Example 2, we have a fair die. We have already seen that if we denote by X the number on the upper face, then X can take the values 1, 2, $\cdots$, 6, each with probability $\frac{1}{6}$. We'll now compute the cumulative distribution function for X. To get an idea of how this is done, let's find $F(2.3)$. (The number 2.3 has been picked a random. There is nothing special about it.) By definition, $F(2.3)$ is the probability that $X \le 2.3$. Now for X to be smaller than or equal to 2.3, we must have either $X = 1$ or $X = 2$. Each of these possibility occurs with probability $\frac{1}{6}$, so

$$F(2.3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

The identical logic and calculation will give $F(x) = \frac{1}{3}$ for any x obeying $2 \le x < 3$. Similarly, if $5 \le x < 6$, then $F(x)$ is the probability that X takes one of the values 1, 2, 3, 4 or 5, which is $\frac{5}{6}$. Here is the full cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{6} & \text{if } 1 \le x < 2 \\ \frac{2}{6} & \text{if } 2 \le x < 3 \\ \frac{3}{6} & \text{if } 3 \le x < 4 \\ \frac{4}{6} & \text{if } 4 \le x < 5 \\ \frac{5}{6} & \text{if } 5 \le x < 6 \\ 1 & \text{if } x \ge 6 \end{cases}$$

and its graph

March 7, 2015
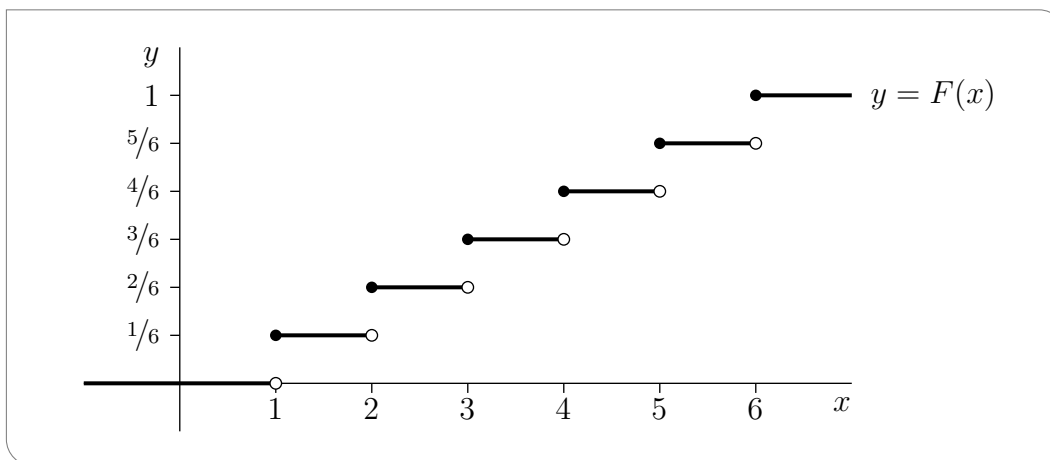
Look carefully at this graph.

- $F(x)$ has jump discontinuities at $x = 1, 2, 3, 4, 5, 6$. These are all of the values that our random variable X can take. $F(x)$ is constant between the jumps.
- In this example, the height of the jump discontinuity at each such x_k is exactly $\frac{1}{6}$, which is the probability $\Pr(X = x_k)$.
- For $x < 1$, the smallest value that X can take, $F(x)$ is identically 0.
- For $x \geq 6$, the largest value that X can take, $F(x)$ is identically 1.
- The graph is like a staircase, moving up to the right.

Example 8

Our observations about the graph of the CDF in the last example were not accidental. For any discrete random variable X, the CDF $F(x)$ of X

- is constant except for jump discontinuities at the values that X can take.
- The height of the jump discontinuity at each such x_k is exactly $\Pr(X = x_k)$.
- For x strictly smaller than the smallest value that X can take, $F(x)$ is identically 0.
- For x larger than, or equal to, the largest value that X can take, $F(x)$ is identically 1.
- The graph is like a staircase, moving up to the right.

Theorem 9 (Properties of the CDF).

If $F(x)$ is a cumulative distribution function, then

- $0 \leq F(x) \leq 1$ for all x.

- $\lim_{x \to -\infty} F(x) = 0$.

- $\lim_{x \to +\infty} F(x) = 1$.

- $F(x)$ is a nondecreasing function of x. This means that if $x < x'$, then $F(x) \leq F(x')$.

   March 7, 2015

Continuous Random Variables

We are now going to concentrate on continuous random variables. For example the height of a randomly chosen person is a continuous random variable. It usually does not make much sense to talk about the probability distribution function $\Pr(X = x)$ of a random variable X that, for example, can take all values between 0 and 1. Because it must satisfy

$$\sum_{0 \leq x \leq 1} P(X = x) = 1 \qquad \text{and} \qquad P(X = x) \geq 0 \text{ for all } x$$

$P(X = x)$ is going to have to be 0 for almost all values of x. In fact, if all x's are equally likely, we will have $P(X = x) = 0$ for *all* values of x — if $P(X = x) = p > 0$ for all $0 \leq x \leq 1$, then $\sum_{0 \leq x \leq 1} P(X = x) = \infty$. So, for continuous random variables, the probability distribution function is not very useful. But the cumulative distribution function still contains all of the probability information about X and so is still useful. That is why we defined it. Recall:

> **Definition 10** (CDF).
>
> The *cumulative distribution function* of *any* random variable X is the function
>
> $$F(x) = \Pr(X \leq x)$$

All cumulative distribution functions obey the properties given in Theorem 9. We can now make a more precise definition of "continuous random variable".

> **Definition 11** (Continuous Random Variable).
>
> A random variable X is a *continuous random variable* if its cumulative distribution function $P(X \leq x)$ is a continuous function of x.

> **Example 12** (Waiting time)
>
> You work in a call centre. Calls come in to the centre at random times. You have just finished dealing with one call. Let X be the length of time before the next call comes in. Typically, X has cumulative distribution
>
> $$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases} \tag{2}$$
>
> Here $\lambda > 0$ is a constant, called the "rate". It is the average number of calls per unit time. (We'll see later, in Example 24, that, for this CDF, the average waiting time for a call to come in is $\frac{1}{\lambda}$. We'll also derive (2) in an optional section at the end of these notes.)

6
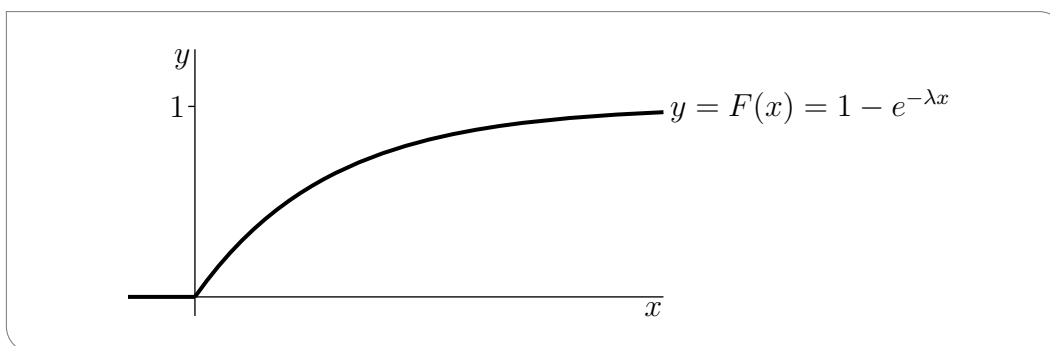
Note that, as required,

- $0 \le F(x) \le 1$ for all x, since, for all $x \ge 0$, we have $0 < e^{-\lambda x} \le 1$.

- $\lim\limits_{x \to -\infty} F(x) = 0$, since $F(x) = 0$ for all $x < 0$.

- $\lim\limits_{x \to +\infty} F(x) = 1 - \lim\limits_{x \to +\infty} e^{-\lambda x} = 1 - 0 = 1$.

- $F(x)$ is a nondecreasing function of x, since, for $x > 0$, $F'(x) = \lambda e^{-\lambda x} > 0$.

Example 12

Once one knows the CDF of a random variable X, one can compute from it lots of other probabilities. For example, if $a < b$, then

$$\Pr(a < X \le b) = \Pr(X \le b) - \Pr(X \le a) = F(b) - F(a) \tag{3}$$

Example 13

Let X be any continuous random variable and fix any real number x. Let $h > 0$. If X happens to obey $X = x$, then it also obeys $x - h < X \le x$, so that $\Pr(X = x) \le \Pr(x - h < X \le x)$. By (3),

$$0 \le \Pr(X = x) \le \Pr(x - h < X \le x) = F(x) - F(x - h)$$

This is true for all $h > 0$. But, since F is continuous, $F(x - h)$ tends to $F(x)$ as h tends to zero and so $F(x) - F(x - h)$ tends to zero as h tends to zero. So taking the limit as h tends to zero of

$$0 \le \Pr(X = x) \le F(x) - F(x - h)$$

gives that

$$0 \le \Pr(X = x) \le 0$$

so that $\Pr(X = x)$ must be exactly zero.

Example 13

For many continuous random variables, $F(x)$ is not only continuous, but is also differentiable. Its derivative is called the probability density function.

7

Definition 14 (PDF).

Let X be a continuous random variable with CDF $F(x)$. If $F(x)$ is differentiable then
$$f(x) = \frac{\mathrm{d}F}{\mathrm{d}x}(x)$$
is called with probability density function (often shortened to PDF) of X.

That is, the CDF $F(x)$ is an antiderivative for the PDF $f(x)$. Then, by the fundamental theorem of calculus,

$$\Pr(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\ dx$$

Since $\lim\limits_{a \to -\infty} F(a) = 0$

$$F(b) = \lim_{a \to -\infty}\big[F(b) - F(a)\big] = \lim_{a \to -\infty}\int_a^b f(x)\ dx = \int_{-\infty}^b f(x)\ dx$$

and since $\lim_{b \to \infty} F(b) = 1$,

$$1 = \lim_{b \to \infty} F(b) = \lim_{b \to \infty}\int_{-\infty}^b f(x)\ dx = \int_{-\infty}^\infty f(x)\ dx$$

We have now seen most of

Theorem 15.

Let $f(x)$ be a probability density function of a random variable X. Then

(a) $f(x) \geq 0$ for all x,

(b) $\int_{-\infty}^\infty f(x)\ dx = 1$,

(c) $F(x) = \Pr(X \leq x) = \int_{-\infty}^x f(t)\ dt$

(d) $\Pr(a < X \leq b) = \int_a^b f(x)\ dx$

Proof. We have already seen all but (a), and it follows from the requirement that $f(x)$ be the derivative of $F(x)$, which is nondecreasing. □

March 7, 2015

Example 16 (Waiting time, again)

Let X be the waiting time random variable of Example 12. Recall that its CDF is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

For $x < 0$ this function has derivative $F'(x) = 0$. For $x > 0$ is has derivative $F'(x) = \lambda e^{-\lambda x}$. At $x = 0$, exactly, its derivative is not defined — look at the graph of $F(x)$ in Example 12. The graph has a kink in it at $x = 0$ and does not have a tangent line there. Set

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \lambda e^{-\lambda x} & \text{if } x > 0 \end{cases}$$
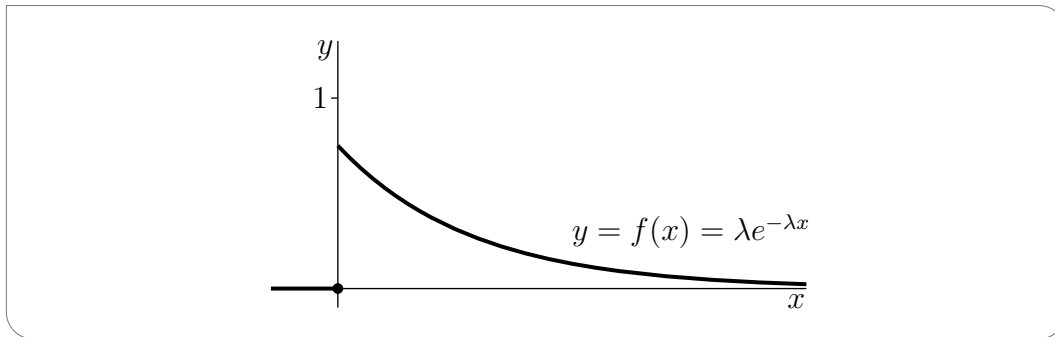
Then $F'(x) = f(x)$ for all x's except $x = 0$, where $F'(x)$ is not defined. It is none–the–less true (because $f(x)$ only has a jump discontinuity at $x = 0$ — it does not go to infinity at $x = 0$) that

$$F(x) = \int_{-\infty}^{x} f(t) \; dt$$

for *all* real numbers x. This is obvious for $x \leq 0$, because both $F(x)$ and $\int_{-\infty}^{x} f(t) \; dt$ are zero when $x \leq 0$. For $x > 0$ it is true because then

$$F(x) = F(x) - F(0) = \int_{0}^{x} f(t) \; dt = \int_{-\infty}^{x} f(t) \; dt$$

Example 16

Random variables like those of Example 16 are fairly common. So we generalize the definition of the "probability density function".

Definition 17 (PDF, generalized).

Let X be a continuous random variable with cumulative distribution function $F(x)$. If a function $f(x)$ obeys

$$F(x) = \int_{-\infty}^{x} f(t) \; dt$$

for all real numbers x, then $f(x)$ is called a probability density function for X.

                   March 7, 2015

It is pretty easy to invent lots of PDF's. Any continuous function (or more generally any function which is continuous except for finitely many jump discontinuities and) which obeys

- $f(x) \geq 0$ for all x and

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

is the PDF for the random variable X with CDF $F(x) = \int_{-\infty}^{x} f(t)\ dt$. It is really easy to invent nonnegative functions. If $\phi(x) \geq 0$ for all x, we can easily construct a PDF out of it by just defining $f(x) = K\phi(x)$, with K chosen so that

$$1 = \int_{-\infty}^{\infty} f(x)\ dx = K \int_{-\infty}^{\infty} \phi(x)\ dx \iff K = \left[\int_{-\infty}^{\infty} \phi(x) \right]^{-1} \tag{4}$$

Example 18

For example, if a random number generator produces each number in the interval $a \leq x \leq b$ with each number being equally likely, then its PDF should be proportional to the "characteristic function"

$$\chi_{[a,b]}(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The funny looking symbol $\chi_{[a,b]}$ is just a standard notation for the function defined by the right hand side of (5). We want the PDF to be proportional to $\chi_{[a,b]}$. That is, we want $f(x) = K\chi_{[a,b]}(x)$ for some constant K. The constant K is determined by the condition that $\int_{-\infty}^{\infty} f(x)\ dx = 1$. Since $\int_{-\infty}^{\infty} K\chi_{[a,b]}(x)\ dx = K \int_a^b dx = K(b-a)$, we must choose $K = \frac{1}{b-a}$, so that the PDF for this random variable is

$$f(x) = \frac{1}{b-a}\chi_{[a,b]}(x)$$

Example 18

Example 19 (Time to failure)

Time to failure, for example of a machine, tends to follow a probability distribution called a "gamma distribution". It has a PDF which is proportional to

$$\phi(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^{\alpha-1} e^{-x/\beta} & \text{if } x \geq 0 \end{cases}$$

with $\alpha > 1$ and $\beta > 0$ being constants. Note that $\phi(x)$ is small if x is small (new machines tend not to fail) or if x is very large (machines don't live forever either). So the PDF of the gamma distribution is

$$f(x) = \frac{1}{K_{\alpha,\beta}} \begin{cases} 0 & \text{if } x < 0 \\ x^{\alpha-1} e^{-x/\beta} & \text{if } x \geq 0 \end{cases}$$

 March 7, 2015

with

$$K_{\alpha,\beta} = \int_0^\infty x^{\alpha-1} e^{-x/\beta} \, dx$$

$$= \beta^\alpha \int_0^\infty u^{\alpha-1} e^{-u} \, du \qquad \text{with } u = {}^x/\beta, \ du = {}^{dx}/\beta$$

$$= \beta^\alpha \Gamma(\alpha) \qquad \text{with } \Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} \, du$$

The function $\Gamma(\alpha)$ is called the gamma function. (That's where the name of the gamma distribution came from.) We can explicitly compute

$$\Gamma(1) = \int_0^\infty e^{-u} \, du = \lim_{R\to\infty} \int_0^R e^{-u} \, du = \lim_{R\to\infty} \left[-e^{-u} \right]_0^R$$
$$= 1$$

and

$$\Gamma(2) = \int_0^\infty u e^{-u} \, du = \lim_{R\to\infty} \int_0^R u e^{-u} \, du$$
$$= \lim_{R\to\infty} \left[-u e^{-u} \Big|_0^R + \int_0^R e^{-u} \, du \right] \quad \text{(by integration by parts with } dv = e^{-u} du, \ v = -e^{-u}\text{)}$$
$$= \lim_{R\to\infty} \left[-u e^{-u} - e^{-u} \right]_0^R$$
$$= 1$$

In fact it is possible to show, by repeated integration by parts, that

$$\Gamma(n) = (n-1)! = 1 \cdot 2 \cdot 3 \cdots (n-1)$$

for all natural numbers[1] n. This was done in Example 11.

Example 19

Expected Value, Variance and Standard Deviation

The probability distribution function gives us extremely detailed information about a random variable. Often it tells us more than we want to to know. For example, suppose that you are charged with deciding how far apart to put the seats in an airplane. You want to know something about the probability distribution of the hip to knee length of your passengers. But you don't want a list of all of their hip to knee lengths. You want a few numbers that give you a decent idea of what the probability distribution looks like. Here is one such number.

[1]It is standard to define $0! = 1$, for notational convenience. Then the formula $\Gamma(n) = (n-1)!$ even works for $n = 1$.

March 7, 2015

Definition 20 (Expected Value).

The *expected value* (or *mean value*, or *expectation*) of a discrete random variable X that can take values x_k, $1 \le k \le n$ is

$$\mathbb{E}(X) = \sum_{k=1}^{n} x_k \Pr(X = x_k)$$

If X is a continuous random variable with PDF $f(x)$, then

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

The expected value gives us the "average value" of X but cannot tell us if the values of X tend to bunch up near $\mathbb{E}(X)$ or if they tend to be very widely scattered about $\mathbb{E}(X)$. For example, for the airplane seating random variable X referred to above, if you only know that $\mathbb{E}(X) = 60$cm and you put your airplane seats 80cm apart it could very well be that

- 99.99% of your passengers can sit down

which is probably OK. But it also could be that

- only 60% of your passengers can sit down.

That would probably not be OK. Here are two related measures of the scatter of a probability distribution.

Definition 21 (Variance, Standard Deviation).

The *variance*, $\mathrm{Var}(X)$, and *standard deviation*, $\sigma(X)$, of a discrete random variable X that can take values x_k, $1 \le k \le n$ are

$$\mathrm{Var}(X) = \sum_{k=1}^{n} \big(x_k - \mathbb{E}(X)\big)^2 \Pr(X = x_k)$$
$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

If X is a continuous random variable with PDF $f(x)$, then

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} \big(x - \mathbb{E}(X)\big)^2 f(x)\ dx$$
$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

 March 7, 2015

Theorem 22 (Another Formula for Variance).

If X is a continuous random variable with PDF $f(x)$, then

$$\text{Var}(X) = \mathbb{E}(X^2) - \big[\mathbb{E}(X)\big]^2$$

Here, the definition of $\mathbb{E}(X^2)$ is

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\, dx$$

Proof. To save writing, set

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx \tag{6}$$

Using this notation, the definition of the variance is

$$\int_{-\infty}^{\infty} (x - \mu)^2\, f(x)\, dx = \int_{-\infty}^{\infty} x^2\, f(x)\, dx - 2\mu \int_{-\infty}^{\infty} x\, f(x)\, dx + \mu^2 \int_{-\infty}^{\infty} f(x)\, dx$$

$$= \mathbb{E}(X^2) - 2\mu^2 + \mu^2$$

$$= \mathbb{E}(X^2) - \mu^2$$

as desired. To get from the first line to the second line, we used (6) and $\int_{-\infty}^{\infty} f(x)\, dx = 1$, since $f(x)$ is a PDF. $\qquad \square$

Example 23 (Beta distribution)

The *beta distribution* is the family of probability distributions with PDF proportional to

$$\phi(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

Here α and β are constants. The beta distribution is used to model continuous random variables that are restricted to finite intervals. For example it is used for time allocation problems in project management.

Let X be a random variable with the $\alpha = \beta = 2$ beta distribution. We'll start by computing the PDF, CDF, expectation value, variance, and standard deviation of X.
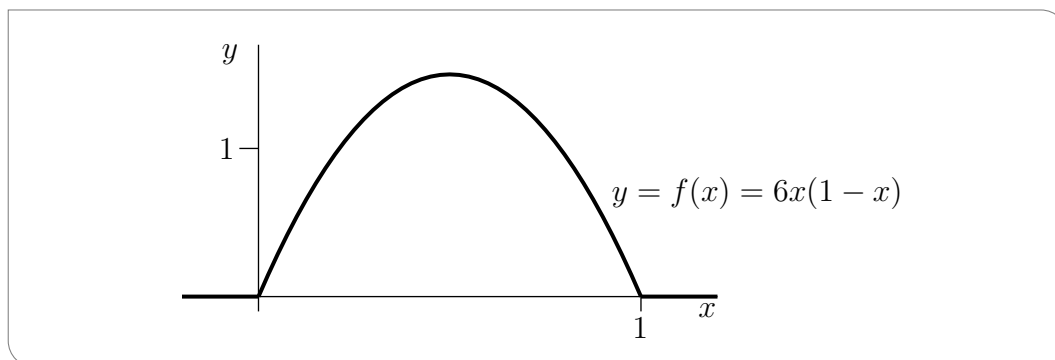
(a) *PDF:* By (4), the PDF of X is

$$f(x) = K \begin{cases} x(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

with

$$\frac{1}{K} = \int_{-\infty}^{\infty} \phi(x)\, dx = \int_0^1 x(1-x)\, dx = \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{6}$$

13

So the PDF is

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$y = f(x) = 6x(1-x)$

(b) *CDF:* Recall that the CDF is $F(x) = \int_{-\infty}^{x} f(t)\, dt$. Because $f(t) = 0$ for all $t \le 0$, $F(x)$ is zero for $x \le 0$. For $0 \le x \le 1$

$$F(x) = \int_0^x 6t(1-t)\, dt = 6\left[\frac{t^2}{2} - \frac{t^3}{3}\right]_0^x = 3x^2 - 2x^3$$

As $f(t) = 0$ for all $t \ge 1$ too, when $x \ge 1$

$$F(x) = \int_0^1 6t(1-t)\, dt = \left[3x^2 - 2x^3\right]_{x=1} = 1$$

So the CDF is

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 3x^2 - 2x^3 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

(c) *Expectation:* The expected value, or mean, of X is

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 6x^2(1-x)\, dx = 6\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1$$
$$= \frac{1}{2}$$

(d) *Variance:* To compute the variance of X, we first compute

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^1 6x^3(1-x)\, dx = 6\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1$$
$$= \frac{3}{10}$$

Using the formula for the variance given in Theorem 22,

$$\text{Var}(X) = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2 = \frac{3}{10} - \left[\frac{1}{2}\right]^2 = \frac{1}{20}$$

14 March 7, 2015

(e) *Standard deviation:* The standard deviation

$$\sigma(X) = \sqrt{\text{Var}(X)} = \frac{1}{\sqrt{20}} = 0.2236$$

to four decimal places.

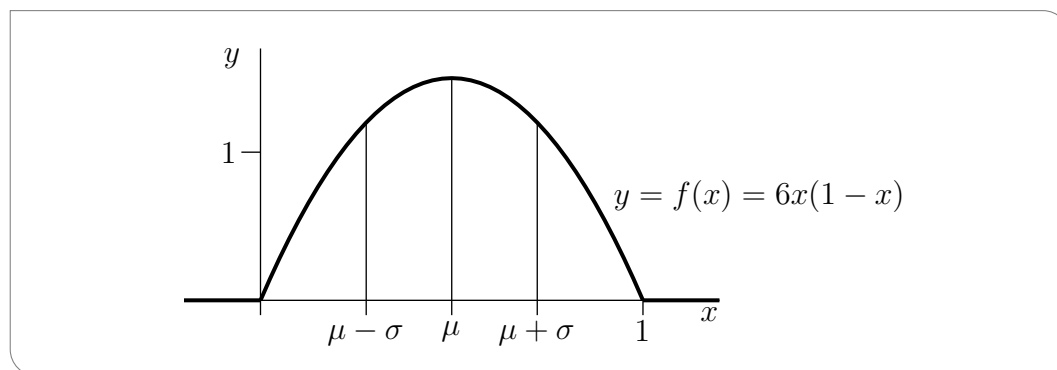(f) *Some probabilities:* If $0 \le a \le b \le 1$, the probability

$$\Pr(a \le x \le b) = \int_a^b 6x(1-x)\ dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_a^b = 3(b^2 - a^2) - 2(b^3 - a^3)$$
$$= b^2(3 - 2b) - a^2(3 - 2a)$$

In particular, the probability of being within one standard deviation of the mean is, with $\mu = 0.5$ and $\sigma = 0.2236$,

$$\Pr(\mu - \sigma \le X \le \mu + \sigma) = \Pr(0.2764 \le X \le 0.7236)$$
$$= 0.7236^2(3 - 2 \times 0.7236) - 0.2764^2(3 - 2 \times 0.2764)$$
$$= 0.6261$$

to four decimal places.

The probability of being within two standard deviations of the mean is, with $\mu = 0.5$ and $2\sigma = 0.4472$,

$$\Pr(\mu - 2\sigma \le X \le \mu + 2\sigma) = \Pr(0.0528 \le X \le 0.9472)$$
$$= 0.9472^2(3 - 2 \times 0.9472) - 0.0528^2(3 - 2 \times 0.0528)$$
$$= 0.9839$$

to four decimal places.

Example 23

March 7, 2015

> **Example 24 (Waiting time, yet again)**

Let X be the waiting time random variable of Example 12. We saw, in Example 16, that its probability density function is

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \lambda e^{-\lambda x} & \text{if } x > 0 \end{cases}$$

We'll now compute its mean and standard deviation. Its mean

$$\mathbb{E}(X) = \lambda \int_0^\infty x e^{-\lambda x} \, dx$$

In Example 19 we saw that

$$K_{\alpha,\beta} = \int_0^\infty x^{\alpha-1} e^{-x/\beta} \, dx = \beta^\alpha \Gamma(\alpha)$$

So

$$\mathbb{E}(X) = \lambda K_{\alpha,\beta} \Big|_{\substack{\alpha=2 \\ \beta=1/\lambda}} = \lambda \left(\frac{1}{\lambda}\right)^2 \Gamma(2)$$

$$= \frac{1}{\lambda}$$

Again, to compute the variance of X, we first compute

$$\mathbb{E}(X^2) = \int_{-\infty}^\infty x^2 f(x) \, dx = \lambda \int_0^1 x^2 e^{-\lambda x} \, dx = \lambda K_{\alpha,\beta} \Big|_{\substack{\alpha=3 \\ \beta=1/\lambda}} = \lambda \left(\frac{1}{\lambda}\right)^3 \Gamma(3)$$

$$= \frac{2}{\lambda^2}$$

Using the formula for the variance given in Theorem 22,

$$\text{Var}(X) = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2 = \frac{2}{\lambda^2} - \left[\frac{1}{\lambda}\right]^2 = \frac{1}{\lambda^2}$$

and the standard deviation is

$$\sigma = \sqrt{\text{Var}(X)} = \frac{1}{\lambda}$$

> **Example 24**

Derivation of the Exponential Waiting Time CDF (Optional)

In Example 12, we pretended that we were working in a call centre, that calls come in to the centre at random times and that we finished dealing with one call at $t = 0$. We denoted by X the length of time before the next call comes in and we assumed that the CDF for X was

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

16

We will now derive this $F(x)$, under some reasonable hypotheses that will be stated shortly.

By definition $F(x) = \Pr(X \le x)$, which is the probability that a call comes in during the time interval $0 < t \le x$. We shall compute, for $x \ge 0$, $G(x) = 1 - F(x)$, which is the probability that no call comes in during the time interval $0 < t \le x$. We shall do so under the assumptions that

- If $a < t \le b$ and $c < t \le d$ are disjoint time intervals, then whether or not there are any calls during $a < t \le b$ has no impact on whether or not there are any calls during $c < t \le d$. Mathematically, this is formulated as

$$\Pr(\text{"no call during } a < t \le b\text{" and "no call during } c < t \le d\text{"})$$
$$= \Pr(\text{"no call during } a < t \le b\text{"}) \; \Pr(\text{"no call during } c < t \le d\text{"})$$

- If $a < b$, then $\Pr(\text{"no call during } a < t \le b\text{"})$ depends only on the length, $b - a$, of the time interval. In particular

$$\Pr(\text{"no call during } a < t \le b\text{"}) = \Pr(\text{"no call during } 0 < t \le b - a\text{"}) = G(b - a)$$

Under these two assumptions, we have that, for any $x, h \ge 0$,

$$
\begin{aligned}
G(x + h) &= \Pr(\text{"no call during } 0 < t \le x + h\text{"}) \\
&= \Pr(\text{"no call during } 0 < t \le x\text{" and "no call during } x < t \le x + h\text{"}) \\
&= \Pr(\text{"no call during } 0 < t \le x\text{"}) \; \Pr(\text{"no call during } x < t \le x + h\text{"}) \\
&= \Pr(\text{"no call during } 0 < t \le x\text{"}) \; \Pr(\text{"no call during } 0 < t \le h\text{"}) \\
&= G(x) \; G(h)
\end{aligned}
$$

We'll also assume that the derivative $G'(0)$ exists. Since $G(0) = 1$ (it's impossible to have a call arrive during $0 < t \le 0$) then the derivative of G at x is

$$
\begin{aligned}
G'(x) &= \lim_{h \to 0} \frac{G(x + h) - G(x)}{h} = \lim_{h \to 0} \frac{G(x)G(h) - G(x)}{h} \\
&= G(x) \lim_{h \to 0} \frac{G(h) - 1}{h} \quad = G(x) \lim_{h \to 0} \frac{G(h) - G(0)}{h} \\
&= G(x)G'(0)
\end{aligned}
$$

Since $G(x)$ is the probability that no call comes in during the time interval $0 < t \le x$, it has to decrease (or stay constant) as x increases. So let's call $G'(0) = -\lambda$, with $\lambda \ge 0$. Then $G(x)$ obeys the differential equation $G'(x) = \lambda G(x)$ and the initial condition $G(0) = 1$. By Example 2 in the notes "Introduction to Differential Equations", with $a = -\lambda$ and $b = 0$, for $x \ge 0$,

$$G(x) = G(0)e^{-\lambda x} = e^{-\lambda x}$$

and

$$F(x) = 1 - G(x) = 1 - e^{-\lambda x}$$

as desired.

 March 7, 2015