Math 345 - Notes
Random Variables
September 24, 2018

# Random Variables

Sometimes, we might want to associate a real value to an outcome. For example, we place a wager on certain outcomes, or we want to count the number of successes in an experiment with many trials. In such an instance, we define a **random variable** $X : \Omega \to \mathbb{R}$ to be a function that takes on values from the set of outcomes and maps into the reals. It is a *variable* because it takes on different values, and it is *random* because it depends on outcomes of random experiments.

**Examples:**

1. You flip a coin. If the coin lands on Heads, you win \$10 and if the coin lands on Tails, you lose \$1. Let $X =$ the net amount of money won in this game. Then we define

$$X(H) = 10 \qquad\qquad X(T) = -1$$

We write $X \in \{-1, 10\}$ to mean that $X$ can take on the values 10 or $-1$. We want to find probabilities for $X$ taking on each value:

$$P(X = 10) = P(\text{Heads}) = 1/2, \qquad P(X = -1) = P(\text{Tails}) = 1/2.$$

A good question to ask is "how much money would I make on this game, on average"? We can say that, since half the time we make \$10, and half the time we lose \$1, we expect to make \$4.5 on average.

2. [extra] Roll a die. Let $X =$the number appearing on the die. Then $X \in \{1, 2, 3, 4, 5, 6\}$ is given by

$$P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = P(X = 6) = 1/6.$$

Note that these probabilities add up to 1.

3. [extra] We toss 3 coins. Let $X$ denote the number of heads. Then $X \in \{0, 1, 2, 3\}$ and

$$\begin{aligned} P(X = 0) &= P(\text{all tails}) = 1/8 \\ P(X = 1) &= P(\text{1 head, 2 tails}) = 3/8 \\ P(X = 2) &= P(\text{2 heads, 1 tail}) = 3/8 \\ P(X = 3) &= P(\text{all heads}) = 1/8 \end{aligned}$$

What is the probability of at most 2 heads?

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}.$$

Another way to find this probability is to observe that

$$P(X \leq 2) = 1 - P(X = 3) = 1 - \frac{1}{8} = \frac{7}{8}.$$

## Discrete random variables

What do we mean by a **discrete** random variable? The random variable takes on "countably many" values. That is, either a finite number like $\{1, 2, 3, 4, 5, 6\}$ or an infinite number of values that we can order, such as $\{0, 1, 2, 3, 4, \ldots\}$. It never takes on all real values in an interval! The examples in the section above are all discrete random variables.
Discrete random variables are fully described by their *probability mass function*, that is, by the list of probabilities $P(X = k)$ for $k$ in the domain of $X$. It satisfies:

$$\sum_{\text{all } k} P(X = k) = 1.$$

Furthermore, we can define the *cumulative distribution function of* $X$ by

$$F(a) = P(X \le a) = \sum_{k \le a} P(X = k).$$

The function $F$ is non-decreasing and non-negative. $F(a)$ encodes the probability accumulated up to the point $a$. Some important discrete random variables are described below.

1. **Bernoulli Random Variables**

   Suppose we run an experiment **ONCE** and it can result in **success or failure**. Suppose the probability of success is $p$ for some constant $p$ satisfying $0 \le p \le 1$. Let

   $$\begin{aligned} X = 1 &\qquad \text{if experiment results in success } \Rightarrow P(X = 1) = p, \\ X = 0 &\qquad \text{if experiment results in failure } \Rightarrow P(X = 0) = 1 - p. \end{aligned}$$

   We call $X$ a Bernoulli random variable with parameter $p$ (which measures success), and denote it by $X = Bernoulli(p)$.

   *Examples:*

   (a) Flip one coin and wish for a Heads. Then getting H is considered a success, so we let $X = 1$ when that occurs. Thus, $P(X = 1) = P(\text{Heads}) = 1/2$.

   (b) Run code and wish to not have any errors. About 70% of the time your code runs without errors, so if we let $X = 1$ for an error-free test, then $P(X = 1) = P(\text{no errors}) = .7$.

2. **Binomial Random Variables:**

   Consider the example: roll four 6-sided dice and let $X$ = number of 6's. Then $X \in \{0, 1, 2, 3, 4\}$, so it is discrete since it takes on 5 values. Find the distribution of $X$.

   $$\begin{aligned} P(X = 0) &= P(\text{no 6's}) = \frac{5^4}{6^4} \\ P(X = 1) &= P(\text{one 6}) = \frac{\binom{4}{1} \times 1 \times 5^3}{6^4} \\ P(X = 2) &= P(\text{two 6's}) = \frac{\binom{4}{2} \times 1^2 \times 5^2}{6^4} \\ P(X = 3) &= P(\text{three 6's}) = \frac{\binom{4}{3} \times 1^3 \times 5^1}{6^4} \\ P(X = 4) &= P(\text{four 6's}) = \frac{\binom{4}{4} \times 1^4}{6^4} \end{aligned}$$

These probabilities come from

$$P(X = k) = (\# \text{ of ways to place k 6's in four spots})(\text{prob. of } 6)^k (\text{prob. of non-6})^{4-k}, \quad \text{for } 0 \le k \le 4$$

**Definition:** The **binomial random variable** with parameters $n$ and $p$ denoted by $Bin(n, p)$ counts the number of successes in $n$ independent trials, with probability of each success being $p$. Its distribution is given for $k \in \{0, 1, 2, \ldots, n\}$ by the formula

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

*Remark:* the reason why we call it a binomial random variable is because its distribution has the combinations $\binom{n}{k}$, which are also called binomials, due to their use in the Binomial Theorem, which states that the coefficients of $(x + y)^n$ are given by $\binom{n}{k}$. More precisely,

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

*Examples of binomial random variables:*

(a) Count the number of heads in 10 coin flips. Then if $X$ is the counter, $X = Bin(10, 0.5)$.

(b) A certain device is functioning with probability 0.9. We test 50 devices and let $X$ be the number that are functioning. Then $X = Bin(50, 0.9)$.

(c) A medical screening test is positive with probability 0.01. If we let $X$ be the number of positive cases out of 1000 people screened for the disease, $X = Bin(1000, 0.01)$.

*Remark:* Note that this agrees with our result for Bernoulli for two reasons. First, $Bernoulli(p) = Bin(1, p)$. Second, if one averages $p$ in one trial, then they should average $np$ in $n$ trials, since $Bin(n, p)$ is just a sequence of $n$ $Bernoulli(p)$ trials.

3. **Poisson Random Variables (optional)**

This random variable is used to model quantities that are given in terms of averages. For example, random variables that obey the Poisson probability law are:

- the number of misprints on a page
- the number of accidents in Redmond today
- the number of people in a community who live to be 100 years old
- the number of $\alpha$-particles discharged in a fixed period of time from a radioactive material
- the number of wrong phone numbers dialed in a day
- the number of DigiPen students fainting in school today

**Definition:** The **Poisson random variable** with parameter $\lambda$ (lambda) and denoted by $Poisson(\lambda)$ counts rare events and has distribution

$$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!},$$

for $k \in \{0, 1, 2, \ldots\}$.

*Examples:*

(a) The number of typographical errors in a textbook from a publisher has an average of $\lambda = 0.5$ errors per page. Find the probability that there is at least one error on the page you are reading.

Let $X$ count the number of errors on the page. We want to find $P(X \geq 1)$. Note that if we knew how many words were on the page, we could use a binomial random variable to model $X$. But we do not know that information. Because the random variable is given in terms of its average, we recognize that in fact $X = Poisson(0.5)$. Then

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \left(e^{-0.5} \cdot \frac{(0.5)^0}{0!}\right) = 1 - e^{-0.5} \approx 0.39.$$

(b) Suppose that earthquakes occur in western US at an average rate of 2 per week. Find the probability that exactly 3 earthquakes occur this week.

Let $X$ = number of earthquakes. Then $X = Poisson(2)$ and $P(X = 3) = e^{-2} \cdot \frac{2^3}{3!} = \frac{4e^{-3}}{3} = 0.18$

4. **Geometric Random Variables**

Suppose we play darts any I hit the target with probability 0.3. I want to play until I **first** hit the target. Let $X$ count how many times I play. Then I can play anywhere between 1 and infinitely many games before hitting the target.

$$\begin{aligned}
P(X = 1) &= P(\text{win 1st}) = 0.3 \\
P(X = 2) &= P(\text{lose 1st, win 2nd}) = (0.7)(0.3) = 0.21 \\
P(X = 3) &= P(\text{lose 1st, 2nd, win 3rd}) = (0.7)(0.7)(0.3) = (.7)^2(0.3) = 1.47 \\
P(X = 100) &= P(\text{lose first 99, win 100th}) = (0.7)^{99}(0.3)
\end{aligned}$$

**Definition:** Let $X$ count the number of trials until the **first** success, with probability of success $p$ in each trial. Then $X$ is a **geometric** random variable with parameter $p$, denoted by $X = Geom(p)$ and whose distribution is defined for $k \in \{1, 2, 3, \dots\}$ by

$$P(X = k) = (1 - p)^{k-1}p.$$

*Remark:* Note that this is an example of a random variable with infinitely many possible values. One can check that all probabilities add up to 1 by computing the infinite sum

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty}(1 - p)^{k-1}p = p\sum_{j=0}^{\infty}(1 - p)^j = p \cdot \frac{1}{1 - (1 - p)} = 1.$$

5. **Negative Binomial Random Variables (optional)**

Continuing with the darts example, where my probability of hitting the target in each trial is 0.3, suppose I play until I hit the target exactly 10 times. Let $X$ count how many times I play. Then I can play anywhere between 10 and infinitely many games before hitting the target exactly 10 times.

$$\begin{aligned}
P(X = 10) &= P(\text{win first 10}) = (0.3)^{10} \\
P(X = 20) &= P(\text{win 20th, win 9 out of the first 19}) = \binom{19}{9}(0.7)^{10}(0.3)^{10}
\end{aligned}$$

The binomial coefficient is used to count all possible ways to win 9 out of the first 19 trials; each success has probability 0.3 and each failure has probability 0.7, so since 10 of each, we have the factors $(0.3)^{10}$ and $(0.7)^{10}$.

**Definition:** Let $X$ count the number of trials until the $r^{th}$ success, with probability of success $p$ in each trial. Then $X$ is a **negative binomial** random variable with parameters $r$ and $p$, which we will denote by $X = NBin(r,p)$ and whose distribution is defined for $k \in \{r, r+1, \dots\}$ by

$$P(X = k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r.$$

## Continuous Random Variables

Continuous random variables take on an uncountable number of values. They are described by their probability density function $f(x)$ which satisfies the properties:

- $f(x) \geq 0$ for $x \in \mathbb{R}$.
- $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$, so the area under the curve equals 1.
- $P(a < X < b) = \displaystyle\int_a^b f(x)\,dx$

Just as for discrete random variables, we define the cumulative distribution function

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx.$$

Remarks:

- $P(X = a) = \displaystyle\int_a^a f(x)\,dx = 0$, for any $a$.
- $f(x) = \dfrac{d}{dx}F(x)$.
- We can interpret $f(x)\,dx$ as the infinitesimal probability of being in a small neighborhood around $x$.

We list here a few common continuous random variables, but there are many more.

1. **Uniform Random Variable**

   **Definition:** Let $X$ denote a point picked at random in the interval $[a,b]$. $X$ has density function

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

   As an example, consider picking a point at random in the interval of reals $[-1,7]$. Here are some probabilities we can compute:

   (a) $P(X = 1) = 0$

   (b) $P(1 \leq X < 4) = \displaystyle\int_1^4 \frac{1}{7-(-1)}\,dx = \frac{3}{8}$.

   (c) $F(3) = P(X \leq 3) = \displaystyle\int_{-1}^3 \frac{1}{7-(-1)}\,dx = \frac{4}{8}$.

2. **Exponential Random Variable (optional)**

   The exponential random variable models waiting times between events, such as time between earthquakes, waiting time at checkout, etc.

   **Definition:** The exponential random variable with parameter $\lambda$ has density function

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

   Example: The time (in hours) required to repair a machine is an exponentially distributed random variable $X$ with parameter $\lambda = 1/2$. We find:

   (a) the probability that a repair time exceeds 2 hours;

$$P(X > 2) = \int_2^{\infty} \frac{1}{2} e^{-x/2}\, dx = \left[-e^{-x/2}\right]_2^{\infty} = e^{-1}.$$

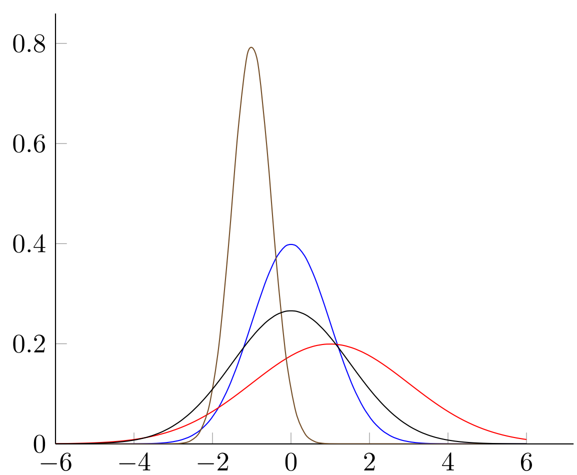
   (b) the conditional probability that a repair takes at least 10 hours, given that its duration exceeds 9 hours.

$$P(X > 10 \,|\, X > 9) = \frac{\int_{10}^{\infty} \frac{1}{2} e^{-x/2}\, dx}{\int_{9}^{\infty} \frac{1}{2} e^{-x/2}\, dx} = \frac{\left[-e^{-x/2}\right]_{10}^{\infty}}{\left[-e^{-x/2}\right]_{9}^{\infty}} = \frac{e^{-10/2}}{e^{-9/2}} = e^{-1/2}.$$

3. **Normal random variables** The most important distribution in statistics and probability is said to be the normal distribution.

   **Definition:** A random variable is said to be **normally distributed** if its distribution has the shape of a bell curve (or normal curve). We write $X = N(\mu, \sigma^2)$ to say that $X$ is a normal random variable with mean $\mu$ (read "mu") and variance $\sigma^2$ (read "sigma" squared). It is described by the **probability density function**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ for } -\infty < x < \infty.$$




   The four normal distributions above have various means and variances, however their shape is the same and the area under each curve is exactly 1.

*Examples of normal random variables:*

(a) the weight and height a person;

(b) velocity of a molecule in gas;

(c) error in measuring a physical quantity;

(d) gestation period of humans.

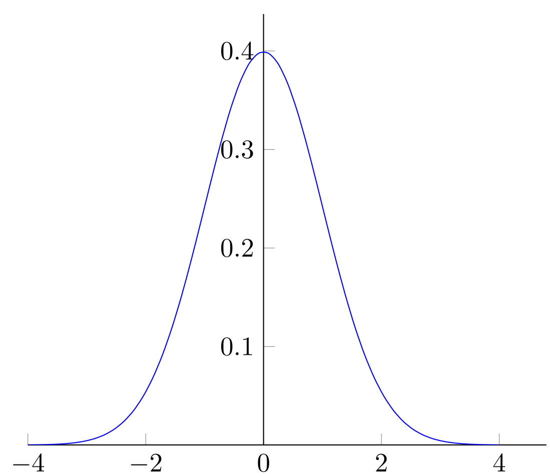
To find the probability

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx,$$

which does not have a nice (closed-form) antiderivative, use your calculator or computational software. It is important to note that the integral above gives the area under the curve $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, from $a$ to $b$.

**Definition:** A normal random variable with $\mu = 0$ and $\sigma = 1$ is called a **standard normal variable**. It has density function $\dfrac{1}{|sqrt 2\pi} e^{-x^2/2}$ and cumulative distribution function

$$\phi(a) = P(X \leq a) = \text{area under the curve to the left of } a.$$




*Properties:*

- The total area under the curve equals 1.
- $P(X \geq a) = 1 - P(X \leq a) = 1 - \phi(a)$
- The area under the curve to the left of $-a$ is equal to the area under the curve to the right of $a$, by symmetry with the $y$-axis, so

$$P(X \leq -a) = P(X \geq a) = 1 - \phi(a).$$

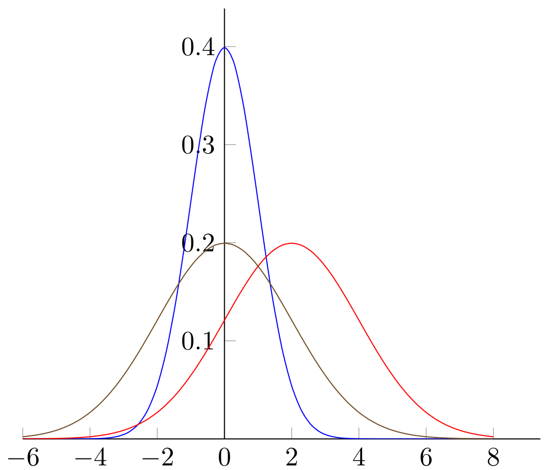
- $P(a \leq X \leq b) = \phi(b) - \phi(a)$.

**Examples: [extra]** Let $X = N(0,1)$

$$
\begin{aligned}
P(X \leq 1) &= \phi(1.00) = .8413 \\
P(X \leq -2) &= \phi(-2) = 1 - \phi(2) = 1 - .9772 = .0128 \\
P(-2 \leq X \leq 1) &= \phi(1) - \phi(-2) = \phi(1) - (1 - \phi(2)) = .8413 - (1 - .9772) = .8185 \\
P(X \geq 0) &= 1 - P(X \leq 0) = 1 - \phi(0) = 1 - .5 = .5 \\
P(X \leq 1.23) &= \phi(1.23) = .8907
\end{aligned}
$$

How does one find probabilities for $N(\mu, \sigma^2)$? Either by using a calculator or by transforming $X = N(\mu, \sigma^2)$ into a standard normal by subtracting the mean and dividing by the standard deviation. That is, let

$$z = \frac{X - \mu}{\sigma}.$$




Visually, standardizing $N(\mu, \sigma^2)$ works as follows. We start with the red curve, which here is $N(2,4)$. Subtracting the mean, shifts to the brown curve, which is $N(0,4)$; it has the same spread, but its mean is now at 0 (where the axis of symmetry is located). Dividing by the standard deviation changes the brown curve into the blue curve which is $N(0,1)$; the brown and blue curves are both centered at 0 (so they have the same mean), but the blue curve has a smaller spread (and the peak is steeper).

**Examples [extra]:** Let $X = N(3,4)$, so $\mu = 3$ and $\sigma = 2$

$$
\begin{aligned}
P(X \leq 5) &= P\left(\frac{X-3}{2} \leq \frac{5-3}{2}\right) = P(z \leq 1) = \phi(1) = .8413 \\
P(X \geq 3) &= P\left(\frac{X-3}{2} \geq \frac{3-3}{2}\right) = P(z \geq 0) = 1 - \phi(0) = .5 \\
P(X \leq 2.5) &= P\left(\frac{X-3}{2} \leq \frac{2.5-3}{2}\right) = P(z \leq -.25) = \phi(-.25) = 1 - \phi(.25) = 1 - .5987 = .4013.
\end{aligned}
$$