Probability II — Random Variables

Prof. Ned Wingreen

MOL 410/510

Random variables

A "random variable" is a function defined on a sample space. A number X is associated with any sample point.

For example, for a coin toss let

$$\begin{aligned} X &= 1 \qquad \text{for heads} \\ X &= 0 \qquad \text{for tails} \end{aligned}$$

The function $P(X = x)$ is the "probability distribution" of the random variable X. Again, for example, if the coin toss is fair,

$$P(X = 1) = P(X = 0) = \frac{1}{2}.$$

We will want to calculate a number of quantities associated with X. We start with the "mean" or "average."

Mean

$$\mu(X) = \mu = \sum_X xP(X = x)$$

The mean is the average value of X after a large number of trials. In our coin example,

$$\mu(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}.$$

Note that different random variables can be defined on the same sample space. In our coin toss example, we could have defined:

$$\begin{aligned} Y &= 1 \qquad \text{for heads} \\ Y &= -1 \qquad \text{for tails} \end{aligned}$$

In that case,

$$\mu(Y) = 1 \cdot P(Y = 1) + (-1) \cdot P(Y = -1) = 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = 0$$

Variance and standard deviation

$$\mathrm{var}(X) = \sum_X (x-\mu)^2 P(X=x)$$

The variance is a useful quality, as is the square root of the variance, the standard deviation:

$$\sigma = \sqrt{\mathrm{var}(X)}, \text{ or} \sigma^2 = \mathrm{var}(X).$$

The variance and standard deviation measure the spread of values of X about the mean. For the two random variables defined for the fair coin tosses,

$$\sigma^2(X) = (1-\frac{1}{2})^2\ \frac{1}{2} + (0-\frac{1}{2})^2\ \frac{1}{2} = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}, \text{ so that } \sigma(X) = \frac{1}{2}$$
$$\sigma^2(Y) = (1)^2\frac{1}{2} + (-1)^2\frac{1}{2} = 1, \text{ so that } \sigma(Y) = 1$$

(The mean and variance of a random variable are also,respectively, called the "1^{st} moment" and the "2^{nd} moment" of the random variable.

Another quantity of interest is the standard error of the mean, $\frac{\sigma}{\sqrt{N}}$. It measures how well we can estimate the mean from a finite sample. We will later derive this quantity and examine more closely its meaning.

Probability distributions

- A random variable that takes on discrete values is a "discrete random variable"
- A random variable that can take on all values in a given range is a "continuous random variable"

Example— Discrete random variable with the binomial distribution

Let X be the number of successes in N independent trials each with probability of success p. Then,

$$P(X=x) = \binom{N}{x} p^x (1-p)^x, \text{ defined on integers } x = 0, 1, \ldots, N.$$

We will now calculate $\mu(X)$ and $\sigma^2(X)$. We could calculate this directly from $P(x)$, but instead we will use these Theorems:

1. The mean of a sum of random variables is the sum of the means.
2. The variances of a sum of uncorrelated random variables is the sum of the variances.

By our assumption, $X = \sum_N X_i$, where

$$X_i = \begin{cases} 1 & \text{with } P(X_i = 1) = p \\ 0 & \text{with } P(X_i = 0) = 1-p \end{cases}$$

By the Theorems,

$$\mu(X) = N\mu(X_i)$$
$$\sigma^2(X) = N\sigma^2(X_i)$$

It is easy to see that for a single trial

$$\mu(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$$
$$\sigma^2(X_i) = (1-p)^2 p + (0-p)^2(1-p) = (1-p)[(1-p)p + p^2] = p(1-p).$$

This finally yields

$$\mu(X) = Np$$
$$\sigma^2(X) = Np(1-p).$$

How does the "width" σ of the binomial distribution compare to its mean?

$$\frac{\sigma}{\mu} = \frac{\sqrt{Np(1-p)}}{Np} = \frac{1}{\sqrt{N}}\sqrt{\frac{1-p}{p}}$$

Relatively speaking, the width gets narrower as $1/\sqrt{N}$, where N is the number of trials.

Example— X with the Poisson distribution

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, \ldots$$

By analogy with the Poisson distribution as the limit of the binomial distribution, the mean of the Poisson distribution is λ and the variance is λ as well. Since $\sigma = \sqrt{\lambda}$,

$$\frac{\sigma}{\mu} = \frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{\mu}}$$

What does the binomial distribution look like for $N = 10, p = 1/2$? A bar graph (only defined on integers). What would the distribution look like for $N = 100$?

Indeed, for most purposes, for large N we can ignore the fact that $P(x)$ is defined only on integers.

Continuous random variables

The probability of achieving a particular value $P(x) = 0$. The probability is only finite when we ask about x in an interval.

Example — Uniform distribution on a circle

This is a continuous fair roulette wheel. After a spin, $x = \theta$. All values of x between 0 and 2π are equally likely.

$$P(x = \pi/4) = 0$$
$$P(0 < x < \pi/4) = 1/8$$
$$P(a < x < b) = \frac{b-a}{2\pi}$$

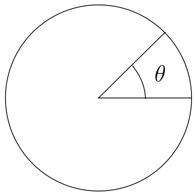

Figure 1: Uniform distribution on a circle

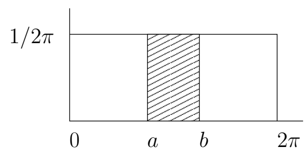

Figure 2: Probability density

Define a probability density:

$$f_X(x) = \frac{1}{2\pi}$$

Then

$$P(a < x < b) = \int_a^b f(x)\,dx = \int_a^b \frac{dx}{2\pi} = \frac{b-a}{2\pi}$$

In general, for a continuous random variable we can define $f(x)$ as

$$P(x < X < x + dx) = f(x)\,dx,$$

so that

$$P(a < x < b) = \int_a^b f(x)\,dx.$$

Then

$$\mu(X) = \int_{-\infty}^{\infty} x f(x) dx = \langle x \rangle = E[x], \text{ the ``expected value'' of } x$$

$$\text{var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \langle (x-\mu)^2 \rangle = E[(x-\mu)^2]$$

In general, if g is a function of x,

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

Note that if X has mean μ and variance σ^2, then

$$Y \equiv X - \mu$$

has mean 0 and variance σ^2. So we can write a random variable as

$$X = \mu + Y,$$

thereby "separating out" the mean.

Normal (Gaussian) distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

The mean is

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

Defining y as $\frac{x-\mu}{\sqrt{2\sigma^2}}$ and noting that $dx = \sqrt{2\sigma^2}dy$, we can substitute:

$$\text{mean} \ = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} (\sqrt{2\sigma^2}y + \mu) e^{-y^2}\, dy$$

The integral of the first term goes to zero, which leaves

$$\frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2}\, dy = \mu$$

The normal distribution is ubiquitous, e.g. heights in a population, measurements, grades — we'll see later that many processes produce normal distributions.

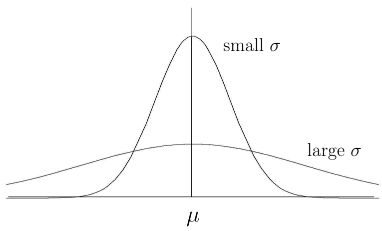

Figure 3: Normal distributions with different σ

If X has a normal distribution with mean μ and standard deviation σ, write

$$X \sim N(\mu, \sigma)$$

We can also define a standardized random variable:

$$z = \frac{X - \mu}{\sigma}$$
$$z \sim N(0, 1)$$

For a normal distribution with mean $\mu = 0$,

$$P(|z| < \sigma) = 0.683$$
$$P(|z| < 2\sigma) = 0.95$$
$$P(|z| < 3\sigma) = 0.997$$

Almost all values of z are within 3 standard deviations of the mean.

(In Matlab `randn` is a standard normal random variable, with $\mu = 0, \sigma^2 = 1$. To get a random variable with mean μ and standard deviation σ, use $\mu + \sigma \cdot$ `randn`.)

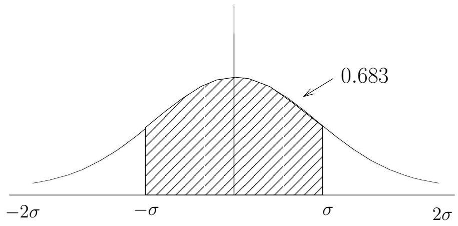

Figure 4: Standardized random variable

Test statistics: z-test and t-test

A test statistic is used to measure the difference between the data and what is expected based on a null hypothesis.

"z" tells how many standard errors an observed mean is from its expected value within a null model. Recall that the standard error of the mean is SE $= \frac{\sigma}{\sqrt{N}}$. So if we accurately know the standard deviation σ of each trial in the null model, we can evaluate $z =$ (observed mean − expected mean)/SE. The integral of the tails of the normal distribution past $|z|$ tell us the probability of observing by chance such a large deviation of the mean from its expected value.

The "t-test"' or "Student's t-test" applies when we don't know the standard deviation σ of the null model, and have to estimate σ from a limited sample (see e.g. *Statistics* by Freedman, Pisani, and Purves, or any other statistics text).

The Law of Large Numbers

Why is the normal distribution so ubiquitous?

The *Law of Large Numbers* – the connection between probability theory and reality:

Random variables X and Y are *independent* if the distribution of X does not depend on the outcome of Y, and vice versa. For example, flipping a coin: the probability of heads on the 21^{st} flip does not depend on the outcome of the 10^{th} flip.

Then, if $X_1, X_2, \ldots, X_n$ are independent identically distributed random variables (discrete or continuous), each with finite mean μ, and if we define

$$S_n = X_1 + X_2 + \cdots + X_n, \text{ and } \varepsilon > 0,$$

then

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

The probability of $\frac{S_n}{n}$, the arithmetic mean, differing from its expected value approaches zero as $n \to \infty$. Note that there is *no* restriction on the variance of X_i.

Central Limit Theorem

This is why the normal distribution is so important.

Let $X_1, X_2, \ldots, X_n$ be identical independent random variables with mean μ and variance σ^2. If we again define S_n as $X_1 + \cdots + X_n$, then

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - \mu n}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-y^2/2}\, dy$$

That is, the random variable $\frac{S_n - \mu n}{\sigma\sqrt{n}} = \frac{S_n/n - \mu}{\sigma/\sqrt{n}}$ is *asymptotically normal*.

This holds for X_i's with *any* distribution, e.g. for the coin toss random variable X, where

$$X = 1 \qquad \text{for heads}$$
$$X = 0 \qquad \text{for tails.}$$

The sum of enough random variables of this type will be normal:

$$S_n \to N(\mu n, \sigma\sqrt{n}) \text{ as } n \to \infty$$

After N coin tosses, the distribution of the number of heads x is the binomial distribution:

$$P(x) = \binom{N}{x} p^x (1-p)^x$$

But the number of heads can also be written as

$$S_N = X_1 + \cdots + X_N,$$

Where the X_i are defined as above. So as $N \to \infty$:

$$S_N(x) = P_{\text{Binomial}}(x) \to N(\mu N, \sigma\sqrt{N}) \text{ as } N \to \infty$$

Therefore, the binomial distribution for large N approaches the normal distribution with the same mean and variance.