- Assume you know what's outcome, sample space, event.
- Conditional probability and independence

  Let A, B be events and $P(B) \neq 0$, then the probability measure of event A after observing the occurrence of event B is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- 2 events A, B are independent iff $P(A \cap B) = P(A)P(B)$

- Bayes's rule:

Bayes's Theorem

From the formula for total probability we immediately obtain one of the most important theorems of elementary probability:

3.5. Bayes's Theorem. Let $A_1, \ldots, A_n \subset S$ be a set of pairwise mutually exclusive events whose union is S and who each have non-zero probability of occurring. Let $B \subset S$ be any event such that $P[B] \neq 0$. Then for any A_k, $k = 1, \ldots, n$,

$$P[A_k \mid B] = \frac{P[B \cap A_k]}{P[B]} = \frac{P[B \mid A_k] \cdot P[A_k]}{\sum_{j=1}^{n} P[B \mid A_j] \cdot P[A_j]}.$$

The theorem is due to the English mathematician ***Thomas Bayes (1701? - 1761)***. Unfortunately, no clearly authentic image of him survives.

- a random variable (RV) X is a function $X: \Omega \to \mathbb{R}$

  (here Ω denotes sample space containing outcomes w)
  - Discrete RV: $X(w)$ can only take finite number of values
  - Continuous RV: $X(w)$ takes on infinite number of possible values

- More generally, we may write

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

to denote the probability that the values of X lie between a and b.

- We assume that a random variable comes with a ***probability density function*** that allows the calculation of probabilities directly, without recourse to the probability space.

- ## Discrete Random Variables

4.2. Definition. Let S be a sample space and Ω a countable subset of $\mathbb{R}$. A ***discrete random variable*** is a map

$$X : S \to \Omega$$

together with a function

$$f_X : \Omega \to \mathbb{R}$$

having the properties that

(i) $f_X(x) \geq 0$ for all $x \in \Omega$ and

(ii) $\sum_{x \in \Omega} f_X(x) = 1$.

The function f_X is called the ***probability density function*** or ***probability distribution*** of X.

We often say that a random variable is given by the pair (X, f_X).

- Expectation of a discrete RV

1.2.4. Definition. Let (X, f_X) be a discrete random variable. Then the expected value of X is

$$\mathrm{E}[X] = \sum_{x \in \Omega} x \cdot f_X(x).$$

provided that the sum (series) on the right converges absolutely.

- Variance of a discrete RV

1.2.8. Definition. Let X be a random variable with expectation $\mathrm{E}[X]$. Then the ***variance*** of X is defined as

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

1.2.9. Notation. For short (and especially in statistics) we write

$$\mathrm{E}[X] = \mu_X = \mu, \qquad \operatorname{Var} X = \sigma_X^2 = \sigma^2.$$

- ## Continuous Random Variables

1.3.1. Definition. Let S be a sample space. A ***continuous random variable*** is a map $X : S \to \mathbb{R}$ together with a function $f_X : \mathbb{R} \to \mathbb{R}$ with the properties that

(i) $f_X \geq 0$ and

(ii) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

no longer the probability at a certain x

The integral of f_X is interpreted as the probability that X assumes values x in a given range, i.e.,

$$P[a \leq X \leq b] = \int_a^b f_X(x)\, dx$$

The function f_X is called the ***probability density function*** (or just density) of the random variable X.

· Expectation and variance of continuous RV

Expectation and Variance

We can define the expectation of a continuous random variable X analogously to that of discrete variables:

$$\mathsf{E}[X] := \int_{\mathbb{R}} x \cdot f_X(x)\, dx$$

It is the possible to prove (using some technical arguments in measure theory) that for any "reasonable" function $\varphi \colon \mathbb{R} \to \mathbb{R}$ we have

$$\mathsf{E}[\varphi \circ X] = \int_{-\infty}^{\infty} \varphi(x) \cdot f_X(x)\, dx,$$

similarly to the discrete case. As before,

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

and all the previously established properties of the expectation and variance continue to hold in the continuous case.

· Bernoulli RV and Binomial RV (Discrete)

Bernoulli Random Variable

Consider an experiment that can result in two possible outcomes, e.g., success or failure, heads or tails, even or odd. Suppose that the probability of success is p, where $0 < p < 1$. Such an experiment is said to be a ***Bernoulli trial***.

4.3. Definition. Let S be a sample space and

$$X \colon S \to \{0, 1\} \subset \mathbb{R}.$$

Let $0 < p < 1$ and define the density function

$$f_X \colon \{0,1\} \to \mathbb{R}, \qquad f_X(x) = \begin{cases} 1-p & \text{for } x = 0 \\ p & \text{for } x = 1. \end{cases}$$

Then X is said to be a ***Bernoulli random variable*** or follow a ***Bernoulli distribution*** with parameter p. We indicate this by writing

$$X \sim \mathsf{Bernoulli}(p)$$

Independent and Identical Trials

More generally, we frequently discuss a sequence of n independent and identical Bernoulli trials. Here,

- ***independent*** means that the outcome of one trial does not influence the outcome of the following trials.
- ***identical*** means that each trial has the same probability of success.

4.4. Example.

- If we flip two fair coins, the two trials are independent and identical.
- If we flip a coin that is fair and another coin that is not fair, the trials are independent but not identical.
- Suppose a box is filled with 10 red balls and 10 black balls. Twice, we draw a ball out of the box but do not replace it. This is a Bernoulli trial where drawing a red ball counts as a "success". The probability of success on the first draw is the same as on the second draw (prove this!). Hence the two trials are identical, but they are clearly not independent. (Since the result of the first draw influences the probability of success in the second draw.)

Counting Successes in a Sequence of Trials

Suppose that we perform a sequence of n independent and identical Bernoulli trials. After recording the results, we define X to be the random variable giving the number of successes in n trials.

To determine the density function of X, we need to find the probability of x successes, where $x = 0, 1, \ldots, n$. Note that a given sequence of results with x successes occurs with probability

$$p^x(1-p)^{n-x}$$

since the probability of success is p and the trials are independent and identical. There are $\binom{n}{x}$ ways to place x successes in n trials, hence there are that many sequences with x successes. Since the sequences are mutually exclusive, their probabilities can be added and we find

$$P[x \text{ successes in } n \text{ trials}] = \binom{n}{x} p^x (1-p)^{n-x}.$$

Binomial Random Variable

4.5. Definition. Let S be a sample space, $n \in \mathbb{N} \setminus \{0\}$, and

$$X : S \to \Omega = \{0, \dots, n\} \subset \mathbb{R}.$$

Let $0 < p < 1$ and define the density function

$$f_X : \Omega \to \mathbb{R}, \qquad f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}. \tag{4.1}$$

Then X is said to be a ***binomial random variable*** with parameters n and p. We indicate this by writing

$$X \sim \mathrm{B}(n, p)$$

Of course, $\mathrm{B}(1, p) = \mathrm{Bernoulli}(p)$.

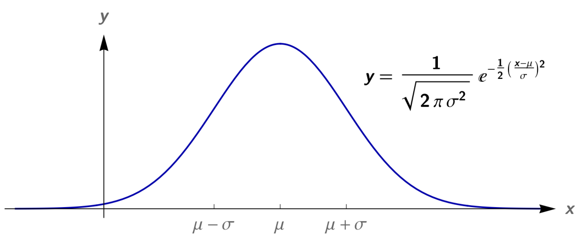
• Normal distribution

Normal (Gauß) Distribution

8.1. Definition. Let $\mu \in \mathbb{R}$, $\sigma > 0$. A continuous random variable (X, f_X) with density

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

is said to follow a normal distribution with parameters μ and σ.



We write

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

whenever a random variable X follows a normal distribution with mean μ and variance σ^2.

8.2. Theorem. Let (X, f_X) be a normally distributed random variable with parameters μ and σ.

(i) $E[X] = \mu$

(ii) $Var[X] = \sigma^2$

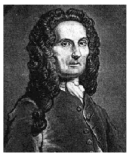
(standard deviation is just the square root of variance)

• 8.8. Theorem. Let X be normally distributed with parameters μ and σ. Then

$$P[-\sigma < X - \mu < \sigma] = 0.68$$
$$P[-2\sigma < X - \mu < 2\sigma] = 0.95$$
$$P[-3\sigma < X - \mu < 3\sigma] = 0.997$$

Hence 68% of the values of a normal random variable lie within one standard deviation of the mean, 95% lie within two standard deviations, and 99.7% lie within three standard deviations. This rule of thumb will be especially important in statistics, where the number of "extraordinary" events needs to be judged.

Approximating the Binomial Distribution



Abraham De Moivre (1667-1754). Portrait. Faber. 1736. File:Abraham de moivre.jpg. (2015, October 30). Wikimedia Commons, the free media repository.

Pierre-Simon de Laplace (1749-1827). Engraving. File:Pierre-Simon-Laplace (1749-1827).jpg. (2017, June 6). Wikimedia Commons, the free media repository.

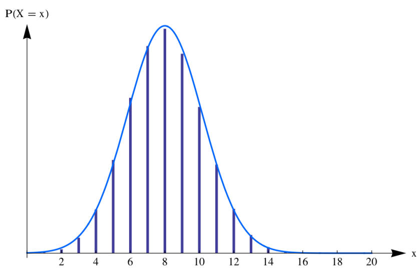
Long before Gauß discovered the normal distribution in 1801, it had been published 60 years earlier, in 1738. De Moivre had wanted to approximate the shape of the binomial distribution, considering the behavior of 3600 coin tosses. In 1810, Laplace proved the general result for $0 < p < 1$.

8.11. Theorem of De Moivre-Laplace. Denote by S_n the number of successes in a sequence of n i.i.d. Bernoulli trials with probability of success $0 < p < 1$. Then

$$\lim_{n\to\infty} P\left[a < \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,dx.$$

Approximating the Binomial Distribution

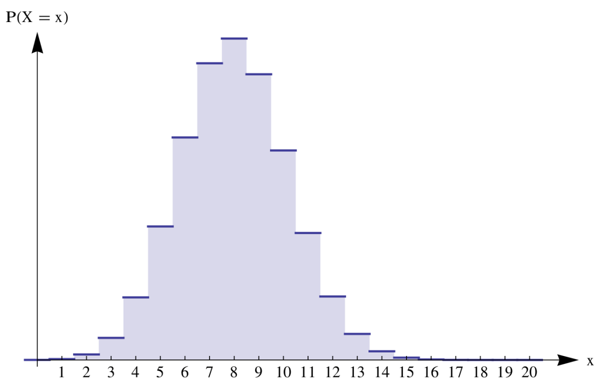
Intuitively, for large n, the binomial distribution with parameters n and p behaves as a normal distribution with mean $\mu = np$ and variance $\sigma^2 = npq$. This is illustrated below for $n = 20$ and $p = 0.4$:



The height of the vertical bars represents the values of $P[X = x]$ according to the binomial distribution, while the density curve of the corresponding normal distribution has been superimposed.

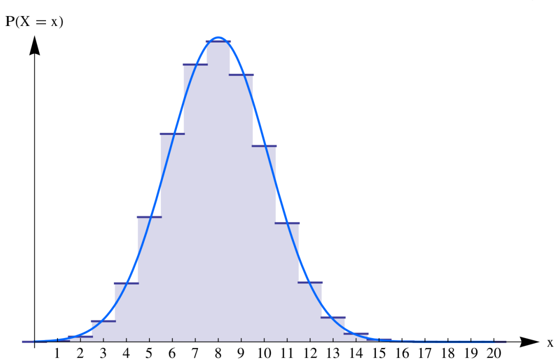
Approximating the Binomial Distribution

We would like to use the normal distribution to approximate the cumulative distribution function of the binomial distribution.



Approximating the Binomial Distribution

It is clear that for each $y = 0, \dots, 20$ the sum over all $x \le y$ corresponds to the area of the bars to the left of y. Superimposing the normal distribution, we see that we can approximate this sum by integrating to $y + 1/2$:



Basically, it says when n, p satisfies some requirement, we can use normal distribution to approximate binomial distribution

(Anything else may not be useful for completing the homework. They've just for fun!)

Approximation and the Half-Unit Correction

Hence, for $y = 0, \dots, n$,

$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi\left(\frac{y + 1/2 - np}{\sqrt{np(1-p)}}\right).$$

here

$$\phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$$

This additional term 1/2 is known as the ***half-unit correction*** for the normal approximation to the cumulative binomial distribution function. It is necessary because in practice we do not have the limit $n \to \infty$ but rather a finite value of n, which may not even be especially large.

This approximation is good if p is close to 1/2 and $n > 10$. Otherwise, we require that

$$np > 5 \qquad \text{if } p \le 1/2 \qquad \text{or} \qquad n(1-p) > 5 \qquad \text{if } p > 1/2.$$

Approximating the Binomial Distribution

8.12. Example. In sampling from a production process that produces items of which 20% are defective, a random sample of 100 items is selected each hour of each production shift. The number of defectives in a sample is denoted by X.

To find, say, $P[X \le 15]$ we might use the normal approximation as follows:

$$\begin{aligned} P[X \le 15] &\approx P\left[Z \le \frac{15 - 100 \cdot 0.2}{\sqrt{100 \cdot 0.2 \cdot 0.8}}\right] = P[Z \le -1.25] \\ &= \Phi(-1.25) = 0.1056 \end{aligned}$$

The half-unit correction would instead give

$$P[X \le 15] \approx P\left[Z \le \frac{15.5 - 20}{4}\right] = 0.130$$

The correct result is $P[X \le 15] = \sum_{k=0}^{15} \binom{100}{k} 0.2^k 0.8^{100-k} = 0.1285$.

Credit to VE401, Probabilistic Methods in Eng. by Horst Hohberger.