**Review: Discrete probability distributions**

*Basic setup of a random experiment:*

1. Sample space: $\Omega$, consists of all elementary outcomes.
2. Set of all events $\mathcal{G}$, which consists of all subsets of $\Omega$.
3. Probability measure: $P : \mathcal{G} \to [0,1]$ which assigns a probability to each event.

*Axioms of Probability:*

1. $P(A) \geq 0$ for all $A \in \mathcal{G}$, $\quad P(\Omega) = 1$.
2. If $A$ and $B$ are disjoint events, then $P(A \cup B) = P(A) + P(B)$.

Important properties:

$$\begin{aligned} P(\bar{A}) = P(\Omega - A) &= 1 - P(A), &(1)\\ P(\emptyset) &= 0, &(2)\\ P(A \cup B) &= P(A) + P(B) - P(A \cap B)\,. &(3) \end{aligned}$$

*Examples:*

1. Roll a die once. Here, $\Omega = \{1,2,3,4,5,6\}$. Each elementary outcome is equally likely, hence $P(\{1\}) = 1/6$, $P(\{2\}) = 1/6$, etc. We can use this to compute the probabilities of events that are not elementary:

$$P(\{1,3,5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{2}\,.$$

2. Roll two dice once. We can represent the sample space in the following form:

| | | | | | |
|---|---|---|---|---|---|
| (1,1) | (1,2) | (1,3) | (1,4) | (1,5) | (1,6) |
| (2,1) | (2,2) | (2,3) | (2,4) | (2,5) | (2,6) |
| (3,1) | (3,2) | (3,3) | (3,4) | (3,5) | (3,6) |
| (4,1) | (4,2) | (4,3) | (4,4) | (4,5) | (4,6) |
| (5,1) | (5,2) | (5,3) | (5,4) | (5,5) | (5,6) |
| (6,1) | (6,2) | (6,3) | (6,4) | (6,5) | (6,6) |

Clearly, we have $6^2 = 36$ possible outcomes which are equally likely. Hence the probability of each elementary outcome is $1/36$.

*Random variables:* A random variable $X$ is a real-valued function on the sample space $\Omega$. If the range of $X$ is finite or countable, $X$ is called discrete, otherwise $X$ is called continuous.
For a discrete random variable $X$ with the range $\{x_1, x_2, ..., x_n\}$, the expectation $\mathbb{E}(X)$ of $X$ (often denoted as $\mu_X$), the variance $\mathrm{Var}(X)$, and the standard deviation $\sigma_X$ are defined as follows:

$$\begin{aligned} \mu_X &\equiv \mathbb{E}(X) = \sum_{k=1}^{n} x_k\, p_k\,, && (4)\\ \mathrm{Var}(X) &= \mathbb{E}\left((X-\mu_X)^2\right) = \sum_{k=1}^{n} (x_k-\mu_X)^2\, p_k\,, && (5)\\ \sigma_X &= \sqrt{\mathrm{Var}(X)}\,. && (6) \end{aligned}$$

Here, $p_k = P(X = x_k)$ is the probability that the random variable takes the value $x_k$. In order to compute this probability, we need to find the set $A \in \mathcal{G}$ that is mapped by $X$ onto $x_k$, or

$$p_k = P(X = x_k) = P(A), \qquad A = X^{-1}(x_k)\,. \tag{7}$$

Note, that as $P(\Omega) = 1$, we have obviously

$$\sum_{k=1}^{n} p_k = 1\,. \tag{8}$$

From the definition of the expectation and the variance, we find directly that, for constants $a$ and $b$ we have

$$\mathbb{E}(aX+b) = a\mathbb{E}(X) + b, \qquad \mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)\,. \tag{9}$$

The proof is left as an exercise. Note that the variance is a quadratic quantity, therefore its scaling factor $a^2$ and not $a$. Moreover, the variance is insensitive to shifts, hence $b$ does not enter on the right-hand side of the second equation.
For the variance, we have the following alternative formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2. \tag{10}$$

Proof: We can see this by direct calculation. Set $\mu = \mathbb{E}(X)$. Then

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right)\\ &= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(X) + \mathbb{E}(\mu^2)\\ &= \mathbb{E}\left(X^2\right) - 2\mu^2 + \mu^2\\ &= \mathbb{E}(X^2) - \mu^2. \end{aligned}$$

*Examples:*

1. Consider again the random experiment of rolling a die once. If we define the random variable $X$ as the number of points of the die, we have $x_k = k$ for $k = 1, 2, ..., 6$. Each $p_k = 1/6$ and we find the $\mu_X = 7/2$. Calculation:

$$\mu_X = \sum_{k=1}^{n} x_k p_k = \frac{1}{6}(1 + 2 + ... + 6) = \frac{7}{2}$$

2. Consider now a second random experiment of rolling two dice once. If we define the random variable $X$ as the sum of the points, the range of $X$ consists of the number from 2 to 12. Now it is (slightly) harder to find the corresponding $p_k$, but if we rewrite the matrix of elementary outcomes again (look at the ascending diagonals)

| | | | | | |
|---|---|---|---|---|---|
| (1,1) | (1,2) | (1,3) | (1,4) | (1,5) | (1,6) |
| (2,1) | (2,2) | (2,3) | (2,4) | (2,5) | (2,6) |
| (3,1) | (3,2) | (3,3) | (3,4) | (3,5) | (3,6) |
| (4,1) | (4,2) | (4,3) | (4,4) | (4,5) | (4,6) |
| (5,1) | (5,2) | (5,3) | (5,4) | (5,5) | (5,6) |
| (6,1) | (6,2) | (6,3) | (6,4) | (6,5) | (6,6) |

we find easily

$$\begin{aligned} P(X = 2) &= P((1,1)) = \frac{1}{36}, \\ P(X = 3) &= P((2,1)) + P((1,2)) = \frac{2}{36}. \end{aligned}$$

*Probability distributions:* From the above examples, it is clear that the function $x_k \to p_k$ captures the essence of the random experiment. In the first case, we have $p_k = 1/6$ for all $x_k$, hence the distribution is uniform, its graph is flat (see Figure 1).

In the second case, $p_k$ increases with $k$ until it reaches its maximum at $x_k = 7$ and then it decreases. Here, the graph is a triangle, see Figure 2.

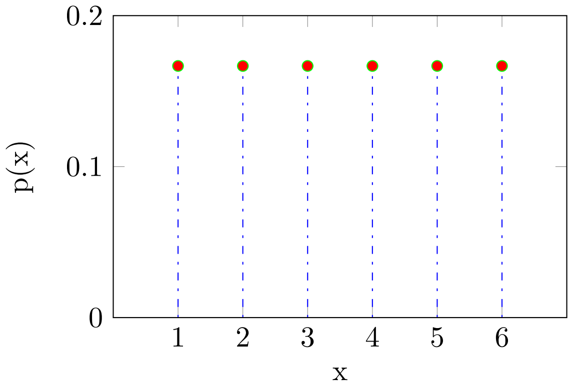



Figure 1: Probability distribution for one die.

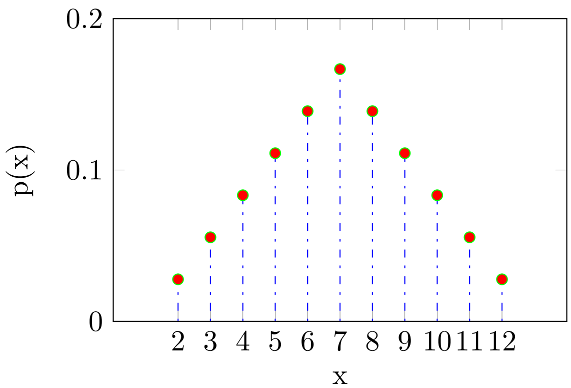



Figure 2: Probability distribution for two dice.

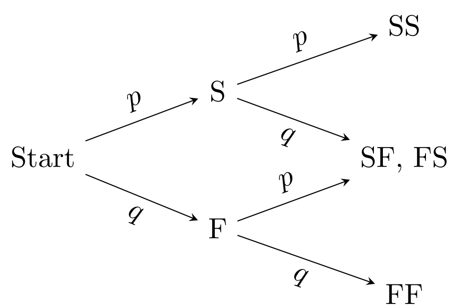
*Binomial distribution:* Consider a random experiment that consists of $n$ independent trials such that

1. each trial has only two outcomes S (success) or F (failure).
2. the probabilities for success $p = P(\text{S})$ and $q = P(\text{F}) = 1 - p$ are the same for all trials.
3. the random variable $X$ counts the number of successes in the $n$ trials

Consider, for example $n = 2$, and $p = q = 1/2$. The possible outcomes of the experiment are $\{SS, SF, FS, FF\}$. The range of $X$ is $\{0, 1, 2\}$ and we find the following probability distribution:

$$
\begin{aligned}
P(X=0) &= P(FF) = \frac{1}{4}\,, \\
P(X=1) &= P(FS) + P(SF) = \frac{1}{2}\,, \\
P(X=2) &= P(SS) = \frac{1}{4}\,.
\end{aligned}
$$

We can represent this experiment in a *tree* diagram:




In the general case, the range of $X$ is $\{0,1,2,...,n\}$. The probability of obtaining $k$ successes during the $n$ trials, can be computed as

$$
P(X=k) = \binom{n}{k} p^k q^{n-k}\,. \tag{11}
$$

Here, $\binom{n}{k}$ is the binomial coefficient, sometimes written as ${}_nC_k$. Hence the binomial distribution is characterized by the two parameters $n$ and $p$ and we write $X \sim B(n,p)$.

Proof: Consider the probability of the particular event $A$ that represents the sequence of obtaining first $k$ successes followed by $n-k$ failures:

$$
A = \underbrace{S...S}_{k}\underbrace{F...F}_{n-k}\,.
$$

Since the probabilities on the branches of the tree multiply due to independence, the probability of this sequence is given by

$$
P\left(\{S...SF...F\}\right) = p^k\, q^{n-k}\,.
$$

Since the random variable $X$ only counts the number of successes, the order in which the successes appear in the sequence is irrelevant: All events with exactly $k$ successes contribute to the total probability $P(X = k)$ and there are exactly $\binom{n}{k}$ such events.

We will later need formulas for the expectation (mean) of $X$ and its variance. For the binomial distribution we have

$$\mathbb{E}(X) = np, \qquad \sigma_X^2 = npq\,. \tag{12}$$

Proof: First remember that, for the sum of two random variables $X_1$ and $X_2$ we have

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)\,, \tag{13}$$

hence for a sum of $n$ random variables we have

$$\mathbb{E}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{E}(X_k)\,. \tag{14}$$

We can represent the binomial random variable $X$ as a sum of Bernoulli random variables

$$X = \sum_{k=1}^{n} X_k\,,$$

where each $X_k$ takes the value 1 in case of success, otherwise the value 0. Clearly, the expected value of $X_k$ is

$$\mathbb{E}(X_k) = p \cdot 1 + q \cdot 0 = p$$

and, therefore, we have $\mathbb{E}(X) = np$. In order to prove the statement about the variance, remember that, for *independent* random variables, the variances add up:

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k)\,. \tag{15}$$

The variance of each $X_k$ can be computed directly:

$$\mathrm{Var}(X_k) = \mathbb{E}(X_k^2) - \mu_k^2 = p - p^2 = p(1-p) = pq\,,$$

which yields $\mathrm{Var}(X) = npq$.

## Review: Continuous probability distributions

In the following we consider a continuous random variable $X$ with a probability density function $p(x)$ on $\mathbb{R}$. One way to think of the probability density function is that the probability that $X$ takes a value in the interval $[x, x + dx)$ is given by

$$P\left(x \leq X < x + dx\right) = p(x)\, dx\,.$$

For continuous probability distributions, the sums in the formulas above become integrals. For instance we have

$$\begin{aligned} 1 &= \int_{\mathbb{R}} p(x)\, dx\,, && (16)\\ \mu_X &= \mathbb{E}(X) = \int_{\mathbb{R}} x\, p(x)\, dx\,, && (17)\\ \sigma_X^2 &= \mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mu_X)^2\, p(x)\, dx\,. && (18) \end{aligned}$$

We also define the cumulative distribution given by $F_X(x) = P(X < x)$ that can be written as

$$F_X(x) = P(X < x) = \int_{-\infty}^{x} p(t)\, dt\,. \tag{19}$$

*Normal distribution:* If the random variable $X$ has the probability density $p$ given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \tag{20}$$

we say that $X$ has a normal distribution with mean $\mu$ and standard deviation $\sigma$, or $X \sim N(\mu, \sigma^2)$.

We can show by direct calculation that indeed

$$\begin{aligned} 1 &= \int_{\mathbb{R}} p(x)\, dx\,, && (21)\\ \mu &= \mathbb{E}(X) = \int_{\mathbb{R}} x\, p(x)\, dx\,, && (22)\\ \sigma^2 &= \mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mu)^2\, p(x)\, dx\,. && (23) \end{aligned}$$

For the proof, we need the following basic calculus result:

$$J = \int_{\mathbb{R}} \mathrm{e}^{-ax^2}\, dx = \sqrt{\frac{\pi}{a}} \tag{24}$$

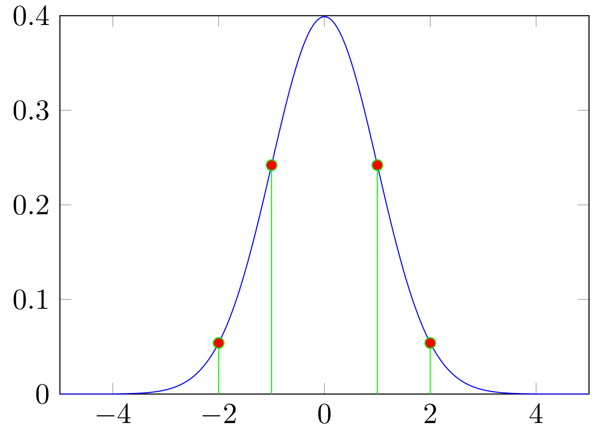



Figure 3: Standard normal distribution with $\mu = 1$, $\sigma = 1$.

Proof of this result: Consider $a = 1$ and instead of $J$ the square $J^2$. Then transform to polar coordinates:

$$\begin{aligned} J^2 &= \left(\int_{\mathbb{R}} \mathrm{e}^{-x^2}\,dx\right)\left(\int_{\mathbb{R}} \mathrm{e}^{-y^2}\,dy\right) \\ &= \int_{\mathbb{R}^2} \mathrm{e}^{-(x^2+y^2)}\,dx\,dy = \int_0^\infty \int_0^{2\pi} \mathrm{e}^{-r^2} r\,d\phi\,dr \\ &= \pi \int_0^\infty 2r\,\mathrm{e}^{-r^2}\,dr = \pi\,. \end{aligned}$$

We can now proceed to prove the equations (21), (22), (23). First, we see immediately that

$$\int_{\mathbb{R}} p(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\,dx = \frac{1}{\sqrt{2\pi}\sigma}\sqrt{2\sigma^2\pi} = 1\,.$$

In order to show (22), we first write

$$\begin{aligned} \int_{\mathbb{R}} x\,p(x)\,dx &= \int_{\mathbb{R}} (x-\mu+\mu)\,p(x)\,dx \\ &= \mu + \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} (x-\mu)\,\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\,dx\,. \end{aligned}$$

Due to symmetry (substitute $\tilde{x} = x - \mu$) the last integral is zero:

$$\int_{\mathbb{R}} (x-\mu)\,\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\,dx = \int_{\mathbb{R}} \tilde{x}\,\mathrm{e}^{-\tilde{x}^2/(2\sigma^2)}\,d\tilde{x} = 0\,.$$

For the last property (23), we need one more integral:

$$\int_{\mathbb{R}} x^2\,\mathrm{e}^{-ax^2}\,dx = \frac{1}{2a}\sqrt{\frac{\pi}{a}}\,. \tag{25}$$

This result also follows directly from what we already know: Use the basic integral (24) to define a function $J(a)$ as

$$J(a) = \int_{\mathbb{R}} \mathrm{e}^{-ax^2}\,dx = \sqrt{\frac{\pi}{a}}$$

and compute $J'(a)$ in two different ways. First, differentiate under the integral and then use the right-hand side of the above equation

$$J'(a) = -\int_{\mathbb{R}} x^2\,\mathrm{e}^{-ax^2}\,dx = \frac{d}{da}\left(\sqrt{\frac{\pi}{a}}\right) = -\frac{1}{2a}\sqrt{\frac{\pi}{a}}\,.$$

Now we can find (23) by direct computation:

$$\begin{aligned}
\int_{\mathbb{R}} (x-\mu)^2\,p(x)\,dx &= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} (x-\mu)^2\,\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\,dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \tilde{x}^2\,\mathrm{e}^{-\tilde{x}^2/(2\sigma^2)}\,d\tilde{x} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\,\frac{2\sigma^2}{2}\sqrt{2\sigma^2\pi} = \sigma^2
\end{aligned}$$

*Central limit theorem*: The mean of a sufficiently large number of iterates of independent random variables, with a well-defined expected value and a well-defined variance, will be approximately normally distributed. In particular, for large $n$, the binomial distribution $B(n,p)$ becomes approximately normal with $N(np, npq)$.

*Moment-generating function:* For a random variable $X$, the moment-generating function $M_X$ is defined as

$$M_X(s) = \mathbb{E}\left(\mathrm{e}^{sX}\right)\,. \tag{26}$$

Clearly, we have $M_X(0) = 1$. More interestingly, we find that, if we know $M_X$, we can compute the mean (and in fact all higher moments) of the random variable $X$ by differentiation. For the mean, this relationship is given by

$$\mu_X = \mathbb{E}(X) = M_X'(0)\,. \tag{27}$$

Proof: From the definition of the moment-generating function, we can use the density $p$ of $X$ to write

$$M'_X(s) = \int_{\mathbb{R}} x \mathrm{e}^{sx} p(x)\, dx$$

If we evaluate this relationship at $s = 0$ we find immediately

$$M'_X(0) = \int_{\mathbb{R}} x p(x)\, dx = \mathbb{E}(X)$$

As an example of how to compute the moment-generating function in a concrete case, take a random variable $Z \sim N(0,1)$. We find by direct calculation:

$$\begin{aligned}
M_Z(s) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{sx - x^2/2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x^2 - 2xs)}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x^2 - 2xs + s^2 - s^2)}\, dx \\
&= \mathrm{e}^{s^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x-s)^2}\, dx = \mathrm{e}^{s^2/2}
\end{aligned}$$

The trick of 'completing the square' used in the third line of the above calculation is very useful in the context of Gaussian integrals.

## Simulating random numbers

In MATLAB, we can generate normally distributed random numbers (e.g. with a standard normal distribution) $Z \sim N(0,1)$ using the command `randn()`. For instance, the command `r=randn(1,1000)` creates a row vector of 1000 random numbers. In order to create histograms, one can use the command `hist()`. The following series of commands compares the histogram of the randomly generated numbers with the theoretical distribution.

```
>> [x,ps] = creategauss(10000);
>> plot(x,ps,x,1/sqrt(2*pi)*exp(-x.^2/2))
```

Here, the function `creategauss()` creates an approximation of the theoretical probability density by a sample of random numbers:

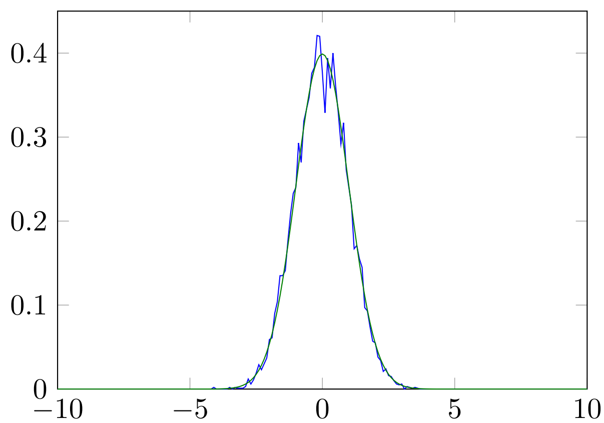



Figure 4: Standard normal distribution from 10000 samples.

```
function [x,ps] = creategauss(n)

% creategauss.m Sampling of a standard normal distribution
%   [x,ps] = creategauss(n);
%   plot(x,ps,x, 1/sqrt(2*pi)*exp(-x.^2/2))

x = [-10:0.1:10]; % create x range
dx = x(2)-x(1);   % here 0.1, of course

r = randn(1,n);   % draw n random numbers
ps = hist(r,x);   % create histogram, centers given by x
ps = ps/sum(ps)*1/dx; % normalize
end
```