Some continuous and discrete distributions

Table of contents

I. Continuous distributions and transformation rules.

A. Standard uniform distribution $U[0,1]$.
B. Uniform distribution $U[a,b]$.
C. Standard normal distribution $N(0,1)$.
D. Normal distribution $N(\mu,\sigma)$. E. Standard exponential distribution.
F. Exponential distribution with mean λ.
G. Standard Gamma distribution $\Gamma(r,1)$.
H. Gamma distribution $\Gamma(r,\lambda)$.

II. Discrete distributions and transformation rules.

A. Bernoulli random variables.
B. Binomial distribution.
C. Poisson distribution.
D. Geometric distribution.
E. Negative binomial distribution.
F. Hypergeometric distribution.

1 Continuous distributions.

Each continuous distribution has a "standard" version and a more general rescaled version. The transformation from one to the other is always of the form $Y = aX + b$, with $a > 0$, and the resulting identities:

$$\begin{aligned} f_Y(y) &= \frac{f_X\left(\frac{y-b}{a}\right)}{a} && (1) \\ F_Y(y) &= F_X\left(\frac{y-b}{a}\right) && (2) \\ E(Y) &= aE(X) + b && (3) \\ Var(Y) &= a^2 Var(X) && (4) \\ M_Y(t) &= e^{bt} M_X(at) && (5) \end{aligned}$$

1.1 Standard uniform $U[0,1]$

This distribution is "pick a random number between 0 and 1".

$$
\begin{aligned}
f_X(x) &= \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \\
F_X(x) &= \begin{cases} 0 & \text{if } x \le 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x \ge 1 \end{cases} \\
E(X) &= 1/2 \\
Var(X) &= 1/12 \\
M_X(t) &= \frac{e^t - 1}{t}
\end{aligned}
$$

1.2 Uniform $U[a,b]$

This distribution is "pick a random number between a and b". To get a random number between a and b, take a random number between 0 and 1, multiply it by $b-a$, and add a. The properties of this random variable are obtained by applying rules (1–5) to the previous subsection.

$$
\begin{aligned}
f_X(x) &= \begin{cases} 1/(b-a) & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases} \\
F_X(x) &= \begin{cases} 0 & \text{if } x \le a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x \ge b \end{cases} \\
E(X) &= (a+b)/2 \\
Var(X) &= (b-a)^2/12 \\
M_X(t) &= \frac{e^{bt} - e^{at}}{t(b-a)}
\end{aligned}
$$

1.3 Standard normal $N(0,1)$

This is the most important distribution in all of probability because of the Central Limit Theorem, which states that the sums (or averages) of a large number of independent random variables is approximately normal, no matter what the original distributions look like. Specifically, if X is a random variable with mean μ and standard deviation σ, and if $X_1, X_2, \ldots$ are independent copies of X, and if $S_n = X_1 + \cdots + X_n$, then for large values of n,

S_n is approximately normal with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$, and $\frac{S_n - n\mu}{\sqrt{n}\sigma}$ is well approximated by the standard normal distribution.

$$\begin{aligned}
f_X(x) &= e^{-x^2/2}/\sqrt{2\pi} \\
F_X(x) & \quad \text{is given in the table at the back of the book.} \\
E(X) &= 0 \\
Var(X) &= 1 \\
M_X(t) &= e^{t^2/2}
\end{aligned}$$

1.4 Normal $N(\mu, \sigma)$

To work with a normal random variable X, convert everything to "Z-scores", where $Z = (X - \mu)/\sigma$. Z is then described by the standard normal distribution, which you can look up in the back of the book. Here are the formulas for X.

$$\begin{aligned}
f_X(x) &= e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2} \\
F_X(x) & \quad \text{is computed from } Z\text{-scores.} \\
E(X) &= \mu \\
Var(X) &= \sigma^2 \\
M_X(t) &= e^{\mu t} e^{\sigma^2 t^2/2}
\end{aligned}$$

1.5 Standard exponential

The exponential distribution describes the time beween successive events in a Poisson process. How long until the next click on my Geiger counter? How long until this lightbulb burns out? How long until the next campaign contribution comes in? A key feature is that it is memoryless: a one-year-old lightbulb has the same change of burning out tomorrow as a brand new lightbulb.

$$\begin{aligned}
f_X(x) &= e^{-x}, \qquad x > 0 \\
F_X(x) &= 1 - e^{-x}, \qquad x > 0
\end{aligned}$$

$$\begin{aligned} E(X) &= 1 \\ Var(X) &= 1 \\ M_X(t) &= 1/(1-t) \end{aligned}$$

1.6 Exponential with mean λ

This is obtained by multiplying a standard exponential by λ. Unfortunately, the letter λ is used differently for Poisson and exponential distributions. If a Poisson distribution has an average rate of r, then the waiting time is exponential with mean $1/r$. When talking about the Poisson distribution we'd be inclined to say "$\lambda = rt$", while when talking about the exponential distribution we'd be inclined to say $\lambda = 1/r$.

$$\begin{aligned} f_X(x) &= \lambda^{-1}e^{-x/\lambda}, \qquad x > 0 \\ F_X(x) &= 1 - e^{-x/\lambda}, \qquad x > 0 \\ E(X) &= \lambda \\ Var(X) &= \lambda^2 \\ M_X(t) &= 1/(1-\lambda t) \end{aligned}$$

1.7 Standard Gamma distribution

The sum of r independent (standard) exponential random variables is called a Gamma random variable. It describes the time you need to wait for r Poisson events to happen (e.g., the time it takes for 10 light bulbs to burn out, for the Geiger counter to record 10 clicks, or for 10 people to send in campaign contributions.) The formula for f_X isn't obvious, and that for F_X is complicated, but the others are directly related to those of the exponential distribution.

$$\begin{aligned} f_X(x) &= x^{r-1}e^{-x}/(r-1)!, \qquad x > 0 \\ F_X(x) &= \text{complicated}, \\ E(X) &= r \\ Var(X) &= r \\ M_X(t) &= (1-t)^{-r} \end{aligned}$$

1.8 Gamma distribution $\Gamma(r, \lambda)$

This is a standard Gamma variable multiplied by λ, or equivalently the sum of r independent exponential variables, each with mean λ

$$
\begin{aligned}
f_X(x) &= \lambda^{-r} x^{r-1} e^{-x/\lambda}/(r-1)!, \qquad x > 0 \\
F_X(x) &= \text{complicated}, \\
E(X) &= \lambda r \\
Var(X) &= \lambda^2 r \\
M_X(t) &= (1 - \lambda t)^{-r}
\end{aligned}
$$

2 Discrete distributions and transformation rules.

The discrete random variables we will consider always take on integer values, so we never rescale them. Also, the cdf $F_X(x)$ is rarely useful, with the notable exception of the geometric distribution. The transformations that are more relevant are those for adding two independent random variables. If $Z = X + Y$, with X and Y independent, then

$$
\begin{aligned}
f_Z(z) &= \sum_x f_X(x) f_Y(z-x) && (6) \\
E(Z) &= E(X) + E(Y) && (7) \\
Var(Z) &= Var(X) + Var(Y) && (8) \\
M_Z(t) &= M_X(t) M_Y(t) && (9)
\end{aligned}
$$

2.1 Bernoulli

A Bernoulli random variable is a variable that can only take on the values 0 and 1. We let p be the probability of 1, and $1 - p$ the probability of 0. This example is easy to analyze, and MANY interesting random variables can be built from this simple building block.

$$
f_X(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases}
$$

$$
\begin{aligned}
E(X) &= p \\
Var(X) &= p(1-p) \\
M_X(t) &= 1 + p(e^t - 1)
\end{aligned}
$$

2.2 Binomial

A binomial random variable is the sum of n independent Bernoulli random variables, all with the same value of p. The usual application is counting the number of successes in n independent tries at something, where each try has a probability p of success. It also applies to sampling with replacement, and is a very good approximation to sampling (without replacement) from very large populations. When np and $n(1-p)$ are both large (say, 20 or bigger), then the binomial distribution is well approximated by the normal distribution. When n is large (at least 30) and p is small (less than 0.1), then the binomial distribution is approximated well by the Poisson distribution.

$$
\begin{aligned}
f_X(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
E(X) &= np \\
Var(X) &= np(1-p) \\
M_X(t) &= (1 + p(e^t - 1))^n
\end{aligned}
$$

2.3 Poisson

The Poisson distribution (pronounced pwah-SON) is the limit of binomial when n is large and p is small. The correspondence is $\lambda = np$. The Poisson distribution replicates itself, in that the sum of a Poisson(λ) random variable and a (independent!) Poisson(μ) random variable is a Poisson($\lambda+\mu$) random variable. Anything that involves the sum of many, many long-shot events (e.g., number of people hit by lightning in a year, or number of broken bones from playing soccer, or number of clicks on a Geiger counter) will be described by the Poisson distribution.

Closely related to the Poisson distribution is a *Poisson process*. In a Poisson process events accumulate at a certain average rate r. The number of events in "time" t is then given by Poisson(rt). Examples include radioactivity, where you have clicks per unit time, typos (where r is errors/page and

t is the number of pages), industrial defects (r equals the average number of defects per foot of sheet metal and t is the number of feet).

$$\begin{aligned}
f_X(x) &= \lambda^x e^{-\lambda}/x! \\
E(X) &= \lambda \\
Var(X) &= \lambda \\
M_X(t) &= e^{\lambda(e^t-1)}
\end{aligned}$$

2.4 Geometric

The geometric distribution describes the waiting time between successes in a Binomial process. For instance, flip coins and count the turns until the first head, or roll dice and count the turns until you get a “6”. It is very much like the exponential distribution, with λ corresponding to $1/p$, except that the geometric distribution is discrete while the exponential distribution is continuous.

$$\begin{aligned}
f_X(x) &= pq^{x-1}, \qquad x = 1, 2, \ldots, \qquad \text{where } q = 1-p \\
E(X) &= 1/p \\
Var(X) &= q/p^2 \\
M_X(t) &= \frac{pe^t}{1-qe^t}
\end{aligned}$$

2.5 Negative binomial

The sum X of r independent geometric random variables is given by the discrete analog of the Gamma distribution (which describes the sum of r independent exponential random variables).

$$\begin{aligned}
f_X(x) &= \binom{x-1}{r-1} p^r q^{x-r}, \qquad x = r, r+1, \ldots, \qquad \text{where } q = 1-p \\
E(X) &= r/p \\
Var(X) &= rq/p^2 \\
M_X(t) &= \frac{p^r e^{rt}}{(1-qe^t)^r}
\end{aligned}$$

The term "negative binomial distribution" actually refers to $Y = X - r$, and not to X. The data for Y are easily obtained from those of X:

$$\begin{aligned}
f_Y(n) &= \binom{n+r-1}{n} p^r q^n, \qquad n = 1, 2, \ldots, \qquad \text{where } q = 1-p \\
E(Y) &= rq/p \\
Var(Y) &= rq/p^2 \\
M_Y(t) &= \frac{p^r}{(1-qe^t)^r}
\end{aligned}$$

2.6 Hypergeometric distribution

The hypergeometric distribution describes sampling *without* replacement. There are r red and w white balls in an urn (with $N = r + w$), we draw n balls out, and let X be the number of red balls drawn. Of course, it doesn't have to be balls in urns. We could just as well be counting aces dealt from a standard deck (with 4 aces and 48 non-aces), or Libertarians drawn from a sample of voters. A hypergeometric random variable is the sum of n Bernoulli random variables (with X_i indicating the color of the i-th ball drawn). These Bernoulli variables are NOT independent, but their covariances are easy to compute, and this gives us the variance of the hypergeometric variable. When $n \ll N$, the hypergeometric distribution is very close to a Binomial distribution with $p = r/N$.

$$\begin{aligned}
f_X(x) &= \frac{\binom{r}{x}\binom{w}{n-x}}{\binom{N}{n}} \\
E(Y) &= nr/N \\
Var(Y) &= n\frac{r}{N}\frac{w}{N}\frac{N-n}{N-1}
\end{aligned}$$