Common probability distributions

D. Joyce, Clark University

Aug 2006

1   Introduction.

I summarize here some of the more common distributions used in probability and statistics. Some are more important than others, and not all of them are used in all fields.

I've identified four sources of these distributions, although there are more than these.

- the uniform distributions, either discrete UNIFORM(n), or continuous UNIFORM(a, b).

- those that come from the Bernoulli process or sampling with or without replacement: BERNOULLI(p), BINOMIAL(n, p), GEOMETRIC(p), NEGATIVEBINOMIAL(p, r), and HYPERGEOMETRIC(N, M, n).

- those that come from the Poisson process: POISSON(λ, t), EXPONENTIAL(λ), GAMMA(λ, r), and BETA(α, β).

- those related to the Central Limit Theorem: NORMAL(μ, σ^2), CHISQUARED(ν), T(ν), and F(ν_1, ν_2).

For each distribution, I give the name of the distribution along with one or two parameters and indicate whether it is a discrete distribution or a continuous one. Then I describe an example interpretation for a random variable X having that distribution. Each discrete distribution is determined by a probability mass function f which gives the probabilities for the various outcomes, so that $f(x) = P(X{=}x)$, the probability that a random variable X with that distribution takes on the value x. Each continuous distribution is determined by a probability density function f, which, when integrated from a to b gives you the probability $P(a \leq X \leq b)$. Next, I list the mean $\mu = E(X)$ and variance $\sigma^2 = E((X - \mu)^2) = E(X^2) - \mu^2$ for the distribution, and for most of the distributions I include the moment generating function $m(t) = E(X^t)$. Finally, I indicate how some of the distributions may be used.

1.1   Background on gamma and beta functions.

Some of the functions below are described in terms of the gamma and beta functions. These are, in some sense, continuous versions of the factorial function $n!$ and binomial coefficients $\binom{n}{k}$, as described below.

The gamma function, $\Gamma(x)$, is defined for any real number x, except for 0 and negative integers, by the integral
$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt.$$
The Γ function generalizes the factorial function in the sense that when x is a positive integer, $\Gamma(x) = (x-1)!$. This follows from the recursion formula, $\Gamma(x+1) = x\Gamma(x)$, and the fact that $\Gamma(1) = 1$, both of which can be easily proved by methods of calculus.

The values of the Γ function at half integers are important for some of the distributions mentioned below. A bit of clever work shows that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, so by the recursion formula,

$$\Gamma\left(n + \tfrac{1}{2}\right) = \frac{(2n-1)(2n-3)\cdots 3 \cdot 1}{2^n} \sqrt{\pi}.$$

The beta function, $\mathrm{B}(\alpha, \beta)$, is a function of two positive real arguments. Like the Γ function, it is defined by an integral, but it can also be defined in terms of the Γ function

$$\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} \, dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

The beta function is related to binomial coefficients, for when α and β are positive integers,

$$\mathrm{B}(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-2)!} = \binom{\alpha+\beta-2}{\alpha-1}^{-1}.$$

2 The uniform distributions

Uniform distributions come in two kinds, discrete and continuous. They share the property that all possible values are equally likely.

2.1 Discrete uniform distribution.

UNIFORM(n). Discrete.

In general, a discrete uniform random variable X can take any finite set as values, but here I only consider the case when X takes on integer values from 1 to n, where the parameter n is a positive integer. Each value has the same probability, namely $1/n$.

$f(x) = 1/n$, for $x = 1, 2, \ldots, n$.

$\mu = \dfrac{n+1}{2}$. $\sigma^2 = \dfrac{n^2-1}{12}$.

$m(t) = \dfrac{1 - e^{(n+1)t}}{1 - e^t}$.

Example. A typical example is tossing a fair cubic die. $n = 6$, $\mu = 3.5$, and $\sigma^2 = \frac{35}{12}$.

2.2 Continuous uniform distribution.

UNIFORM(a, b). Continuous.

In general, a continuous uniform variable X takes values on a curve, surface, or higher dimensional region, but here I only consider the case when X takes values in an interval $[a, b]$. Being uniform, the probability that X lies in a subinterval is proportional to the length of that subinterval.

$$f(x) = \frac{1}{b-a}, \text{ for } x \in [a, b].$$
$$\mu = \frac{a+b}{2}. \quad \sigma^2 = \frac{(b-a)^2}{12}.$$
$$m(t) = \frac{e^{bt} - e^{at}}{t(b-a)}.$$

Example. An example of an approximately uniform distribution on the interval $[-\frac{1}{2}, \frac{1}{2}]$ is the roundoff error when rounding to the nearest integer. This assumes that the distribution of numbers being rounded off is fairly smooth and stretches over several integers.

Note. Most computer programming languages have a built in pseudorandom number generator which generates numbers X in the unit interval $[0, 1]$. Random number generators for any other distribution can then computed by applying the inverse of the cumulative distribution function for that distribution to X.

3 The Bernoulli process and sampling with or without replacement

A single trial for a Bernoulli process, called a Bernoulli trial, ends with one of two outcomes, one often called success, the other called failure. Success occurs with probability p while failure occurs with probability $1 - p$, usually denoted q.

The Bernoulli process consists of repeated independent Bernoulli trials with the same parameter p. These trials form a random sample from the Bernoulli population.

You can ask various questions about a Bernoulli process, and the answers to these questions have various distributions. If you ask how many successes there will be among n Bernoulli trials, then the answer will have a binomial distribution, BINOMIAL(n, p). The Bernoulli distribution, BERNOULLI(p), simply says whether one trial is a success. If you ask how many trials it will be to get the first success, then the answer will have a geometric distribution, GEOMETRIC(p). If you ask how many trials there will be to get the rth success, then the answer will have a negative binomial distribution, NEGATIVEBINOMIAL(p, r). Given that there are M successes among N trials, if you ask how many of the first n trials are successes, then the answer will have a HYPERGEOMETRIC(N, M, n) distribution.

Sampling with replacement is described below under the binomial distribution, while sampling without replacement is described under the hypergeometric distribution.

3.1 Bernoulli distribution.

BERNOULLI(p). Discrete.

The parameter p is a real number between 0 and 1. The random variable X takes on two values: 1 (success) or 0 (failure). The probability of success, $P(X=1)$, is the parameter p. The symbol q is often used for $1 - p$, the probability of failure, $P(X=0)$.

$f(0) = 1 - p$, $f(1) = p$.
$\mu = p$. $\sigma^2 = p(1-p) = pq$.
$m(t) = pe^t + q$.

Example. If you flip a fair coin, then $p = \frac{1}{2}$, $\mu = p = \frac{1}{2}$, and $\sigma^2 = \frac{1}{4}$.

3.2 Binomial distribution.

BINOMIAL(n, p). Discrete.

Here, X is the sum of n independent Bernoulli trials, each BERNOULLI(p), so $X = x$ means there were x successes among the n trials. (Note that the BERNOULLI(p) = BINOMIAL($1, p$).)

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, for $x = 0, 1, \ldots, n$.
$\mu = np$. $\sigma^2 = np(1-p) = npq$.
$m(t) = (pe^t + q)^n$.

Sampling with replacement. Sampling with replacement occurs when a set of N elements has a subset of M "preferred" elements, and n elements are chosen at random, but the n elements don't have to be distinct. In other words, after one element is chosen, it's put back in the set so it can be chosen again. Selecting a preferred element is success, and that happens with probability $p = M/N$, while selecting a nonpreferred element is failure, and that happens with probability $q = 1 - p = (N - M)/N$. Thus, sampling with replacement is a Bernoulli process.

A bit of history. The first serious development in the theory of probability was in the 1650s when Pascal and Fermat investigated the binomial distribution in the special case $p = \frac{1}{2}$. Pascal published the resulting theory of binomial coefficients and properties of what we now call Pascal's triangle.

In the very early 1700s Jacob Bernoulli extended these results to general values of p.

3.3 Geometric distribution.

GEOMETRIC(p). Discrete.

When independent Bernoulli trials are repeated, each with probability p of success, the number of trials X it takes to get the first success has a geometric distribution.

$f(x) = q^{x-1}p$, for $x = 1, 2, \ldots$.
$\mu = \dfrac{1}{p}$. $\sigma^2 = \dfrac{1-p}{p^2}$.
$m(t) = \dfrac{pe^t}{1 - qe^t}$.

3.4 Negative binomial distribution.

NEGATIVEBINOMIAL(p, r). Discrete.

When independent Bernoulli trials are repeated, each with probability p of success, and X is the trial number when r successes are first achieved, then X has a negative binomial distribution. Note that GEOMETRIC(p) = NEGATIVEBINOMIAL$(p, 1)$.

$f(x) = \binom{x-1}{r-1} p^r q^{x-r}$, for $x = r, r+1, \ldots$.

$\mu = \dfrac{r}{p}$. $\sigma^2 = \dfrac{rq}{p^2}$.

$m(t) = \left(\dfrac{pe^t}{1 - qe^t} \right)^r$.

3.5 Hypergeometric distribution.

HYPERGEOMETRIC(N, M, n). Discrete.

In a Bernoulli process, given that there are M successes among N trials, the number X of successes among the first n trials has a HYPERGEOMETRIC(N, M, n) distribution.

$f(x) = \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$, for $x = 0, 1, \ldots, n$.

$\mu = np$. $\sigma^2 = \binom{N-n}{N-1} npq$.

Sampling without replacement. Sampling with replacement was mentioned above in the section on the binomial distribution. Sampling without replacement is similar, but once an element is selected from a set, it is taken out of the set so that it can't be selected again.

Formally, we have a set of N elements with a subset of M "preferred" elements, and n distinct elements among the N elements are to be chosen at random. The symbol p is often used for the probability of choosing a preferred element, that is, $p = M/N$, and $q = 1 - p$. The hypergeometric distribution, described below, answers the question: of the n chosen elements, how many are preferred?

When n is a small fraction of N, then sampling without replacement is almost the same as sampling with replacement, and the hypergeometric distribution is almost a binomial distribution.

Many surveys have questions with only two responses, such as yes/no, for/against, or candidate A/B. These are actually sampling without replacement because the same person won't be asked to respond to the survey twice. But since only a small portion of the population will respond, the analysis of surveys can be treated as sampling with replacement rather than sampling without replacement (which it actually is).

4 The Poisson process

A Poisson process is the continuous version of a Bernoulli process. In a Bernoulli process, time is discrete, and at each time unit there is a certain probability p that success occurs, the same probability at any given time, and the events at one time instant are independent of the events at other time instants.

In a Poisson process, time is continuous, and there is a certain rate λ of events occurring per unit time that is the same for any time interval, and events occur independently of each other. Whereas in a Bernoulli process either no or one event occurs in a unit time interval, in a Poisson process any nonnegative whole number of events can occur in unit time.

As in a Bernoulli process, you can ask various questions about a Poisson process, and the answers will have various distributions. If you ask how many events occur in an interval of length t, then the answer will have a Poisson distribution, $\text{POISSON}(\lambda t)$. If you ask how long until the first event occurs, then the answer will have an exponential distribution, $\text{EXPONENTIAL}(\lambda)$. If you ask how long until the rth event, then the answer will have a gamma distribution, $\text{GAMMA}(\lambda, r)$. If there are $\alpha + \beta$ events in a given time interval, if you ask what fraction of the interval it takes until the αth event occurs, then the answer will have a $\text{BETA}(\alpha, \beta)$ distribution.

4.1 Poisson distribution.

$\text{POISSON}(\lambda t)$. Discrete.

When events occur uniformly at random over time at a rate of λ events per unit time, then the random variable X giving the number of events in a time interval of length t has a Poisson distribution.

Often the Poisson distribution is given just with the parameter λ, but I find it useful to incorporate lengths of time t other than requiring that $t = 1$.

$f(x) = \frac{1}{x!}(\lambda t)^x e^{-\lambda t}$, for $x = 0, 1, \dots$.
$\mu = \lambda t$. $\sigma^2 = \lambda t$.
$m(s) = e^{\lambda t(e^s - 1)}$.

4.2 Exponential distribution.

$\text{EXPONENTIAL}(\lambda)$. Continuous.

When events occur uniformly at random over time at a rate of λ events per unit time, then the random variable X giving the time to the first event has an exponential distribution.

$f(x) = \lambda e^{-\lambda x}$, for $x \in [0, \infty)$.
$\mu = 1/\lambda$. $\sigma^2 = 1/\lambda^2$.
$m(t) = (1 - t/\lambda)^{-1}$.

4.3 Gamma distribution.

$\text{GAMMA}(\lambda, r)$ or $\text{GAMMA}(\alpha, \beta)$. Continuous.

In the same Poisson process for the exponential distribution, the gamma distribution gives the time to the rth event. Thus, $\text{EXPONENTIAL}(\lambda) = \text{GAMMA}(\lambda, 1)$.

The gamma distribution also has applications when r is not an integer. For that generality the factorial function is replaced by the gamma function, $\Gamma(x)$, described above.

There is an alternate parameterization $\text{GAMMA}(\alpha, \beta)$ of the family of gamma distributions. The connection is $\alpha = r$, and $\beta = 1/\lambda$ which is the expected time to the first event in a Poisson process.

$$f(x) = \frac{1}{\Gamma(r)}\lambda^r x^{r-1} e^{-\lambda x} = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \text{ for } x \in [0, \infty).$$
$$\mu = r/\lambda = \alpha\beta. \; \sigma^2 = r/\lambda^2 = \alpha\beta^2.$$
$$m(t) = (1 - t/\lambda)^{-r} = (1 - \beta t)^{-\alpha}.$$

Application to Bayesian statistics. Gamma distributions are used in Bayesian statistics as conjugate priors for the distributions in the Poisson process. In GAMMA(α, β), α counts the number of occurrences observed while β keeps track of the elapsed time.

4.4 Beta distribution.

BETA(α, β). Continuous.

In a Poisson process, if $\alpha + \beta$ events occur in a time interval, then the fraction of that interval until the αth event occurs has a BETA(α, β) distribution.

Note that BETA$(1, 1) = $ UNIFORM$(0, 1)$.

In the following formula, B(α, β) is the beta function described above.

$$f(x) = \frac{1}{\text{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \text{ for } 0 \le x \le 1.$$
$$\mu = \frac{\alpha}{\alpha + \beta}. \; \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Application to Bayesian statistics. Beta distributions are used in Bayesian statistics as conjugate priors for the distributions in the Bernoulli process. Indeed, that's just what Thomas Bayes (1702–1761) did. In BETA(α, β), α counts the number of successes observed while β keeps track of the failures observed.

5 Distributions related to the central limit theorem

The Central Limit Theorem says sample means and sample sums approach normal distributions as the sample size approaches infinity.

5.1 Normal distribution.

NORMAL(μ, σ^2). Continuous.

The normal distribution, also called the Gaussian distribution, is ubiquitous in probability and statistics.

The parameters μ and σ^2 are real numbers, σ^2 being positive with positive square root σ.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \text{ for } x \in \mathbf{R}.$$
$$\mu = \mu. \; \sigma^2 = \sigma^2.$$
$$m(t) = \exp(\mu t + t^2\sigma^2/2).$$

The standard normal distribution has $\mu = 0$ and $\sigma^2 = 1$. Thus, its density function is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Applications. Normal distributions are used in statistics to make inferences about the population mean when the sample size n is large.

Application to Bayesian statistics. Normal distributions are used in Bayesian statistics as conjugate priors for the family of normal distributions with a known variance.

5.2 χ^2-distribution.

CHISQUARED(ν). Continuous.

The parameter ν, the number of "degrees of freedom," is a positive integer. I generally use this Greek letter nu for degrees of freedom. This is the distribution for the sum of the squares of ν independent standard normal distributions.

A χ^2-distribution is actually a special case of a gamma distribution with a fractional value for r. CHISQUARED(ν) = GAMMA(λ, r) where $\lambda = \frac{1}{2}$ and $r = \frac{\nu}{2}$.

$$f(x) = \frac{x^{\nu/2-1}e^{x/2}}{2^{\nu/2}\Gamma(\nu/2)}, \text{ for } x \geq 0.$$
$$\mu = \nu. \ \sigma^2 = 2\nu.$$
$$m(t) = (1 - 2t)^{\nu/2}.$$

Applications. χ^2-distributions are used in statistics to make inferences on the population variance when the population is assumed to be normally distributed.

5.3 Student's T-distribution.

T(ν). Continuous.

If Y is NORMAL(0, 1) and Z is CHISQUARED(ν) and independent of Y, then $X = Y/\sqrt{Z/\nu}$ has a T-distribution with ν degrees of freedom.

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})\,(1 + x^2/\nu)^{(\nu+1)/2}}, \text{ for } x \in \mathbf{R}.$$
$$\mu = 0. \ \sigma^2 = \frac{\nu}{\nu-2} \text{ for } \nu > 2.$$

Applications. T-distributions are used in statistics to make inferences on the population variance when the population is assumed to be normally distributed, especially when the population is small.

5.4 Snedecor-Fisher's F-distribution.

F(ν_1, ν_2). Continuous.

If Y and Z are independent χ^2-random variables with ν_1 and ν_2 degrees of freedom, respectively, then $X = \dfrac{Y/\nu_1}{Z/\nu_2}$ has an F-distribution with (ν_1, ν_2) degrees of freedom.

Note that if X is T(ν), then X^2 is F(1, ν).

$$f(x) = \frac{1}{B(\frac{\nu_1}{2}, \frac{\nu_2}{2})} \frac{(\frac{\nu_1}{\nu_2})^{\nu_1/2}x^{\nu_1/2-1}}{(1 + \frac{\nu_1}{\nu_2}x)^{(\nu_1+\nu_2)/2}} \text{ for } x > 0.$$

$$\mu = \frac{\nu_2}{\nu_2 - 2} \text{ when } \nu_2 > 2. \quad \sigma^2 = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \text{ when } \nu_2 > 4.$$

Applications. F-distributions are used in statistics when comparing variances of two populations.

Table of Discrete and Continuous distributions

Distribution	Type	Mass/density function $f(x)$	Mean μ	Variance σ^2
UNIFORM(n)	D	$1/n$, for $x = 1, 2, \ldots, n$	$(n+1)/2$	$(n^2 - 1)/12$
UNIFORM(a,b)	C	$\frac{1}{b-a}$, for $x \in [a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
BERNOULLI(p)	D	$f(0) = 1 - p,\ f(1) = p$	p	$p(1-p)$
BINOMIAL(n,p)	D	$\binom{n}{x} p^x (1-p)^{n-x}$, for $x = 0, 1, \ldots, n$	np	npq
GEOMETRIC(p)	D	$q^{x-1}p$, for $x = 1, 2, \ldots$	$1/p$	$(1-p)/p^2$
NEGATIVEBINOMIAL(p,r)	D	$\binom{x-1}{r-1} p^r q^{x-r}$, for $x = r, r+1, \ldots$	r/p	$r(1-p)/p^2$
HYPERGEOMETRIC(N,M,n)	D	$\dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$, for $x = 0, 1, \ldots, n$	np	$np(1-p)$
POISSON(λt)	D	$\frac{1}{x!}(\lambda t)^x e^{-\lambda t}$, for $x = 0, 1, \ldots$	λt	λt
EXPONENTIAL(λ)	C	$\lambda e^{-\lambda x}$, for $x \in [0, \infty)$	$1/\lambda$	$1/\lambda^2$
GAMMA(λ, r)	C	$\frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x}$	r/λ	r/λ^2
GAMMA(α, β)		$= \dfrac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$, for $x \in [0, \infty)$	$= \alpha\beta$	$= a\beta^2$
BETA(α, β)	C	$\frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$, for $0 \le x \le 1$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
NORMAL(μ, σ^2)	C	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, for $x \in \mathbf{R}$	μ	σ^2
CHISQUARED(ν)	C	$\frac{x^{\nu/2-1} e^{x/2}}{2^{\nu/2} \Gamma(\nu/2)}$, for $x \ge 0$	ν	2ν
T(ν)	C	$\dfrac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\, \Gamma(\frac{\nu}{2})\, (1 + x^2/\nu)^{(\nu+1)/2}}$ for $x \in \mathbf{R}$	0	$\nu/(\nu - 2)$
F(ν_1, ν_2)	C	$\dfrac{1}{\mathrm{B}(\frac{\nu_1}{2}, \frac{\nu_2}{2})} \dfrac{(\frac{\nu_1}{\nu_2})^{\nu_1/2} x^{\nu_1/2-1}}{(1 + \frac{\nu_1}{\nu_2}x)^{(\nu_1+\nu_2)/2}}$ for > 0	$\dfrac{\nu_2}{\nu_2 - 2}$	$\frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$