RANDOM VARIABLES

A random variable is **discrete** if the set of outcomes is either finite in number or countably infinite. the random variable is **continuous** if the set of outcomes is infinitely divisible and, hence, not countable.

RANDOM VARIABLES

- PROBABILITY DISTRIBUTIONS

A listing of the values x taken by a random variable X and their associated probabilities is a **probability distribution**, $f(x)$. For a discrete random variable,

$$f(x) = \text{Prob}(X = x).$$

The **axioms of probability** require that

1. $0 \leq \text{Prob}(X = x) \leq 1.$
2. $\sum_x f(x) = 1.$

The **probability density function (pdf)**, $f(x)$, is defined so that $f(x) \geq 0$ and

1. $\text{Prob}(a \leq x \leq b) = \int_a^b f(x)\, dx \geq 0.$
2. $\int_{-\infty}^{+\infty} f(x)\, dx = 1.$

RANDOM VARIABLES

- CUMULATIVE DISTRIBUTION FUNCTION

For a discrete random variable,

$$F(x) = \sum_{X \leq x} f(X) = \text{Prob}(X \leq x).$$

For a continuous random variable,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt,$$

EXPECTATIONS OF A RANDOM VARIABLE

The ***mean****, or* ***expected value****, of a random variable is*

$$E[x] = \begin{cases} \sum_x xf(x) & \text{if } x \text{ is discrete,} \\ \int_x xf(x)\,dx & \text{if } x \text{ is continuous.} \end{cases}$$

If $g(x) = a + bx$ for constants a and b, then

$$E[a + bx] = a + bE[x].$$

An important case is the expected value of a constant a, which is just a.

EXPECTATIONS OF A RANDOM VARIABLE

*The **variance** of a random variable is*

$$\mathrm{Var}[x] = E[(x-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 f(x) & \text{if } x \text{ is discrete,} \\ \int_x (x-\mu)^2 f(x)\,dx & \text{if } x \text{ is continuous.} \end{cases}$$

Computation of the variance is simplified by using the following important result:

$$\mathrm{Var}[x] = E[x^2] - \mu^2.$$

$$\mathrm{Var}[a+bx] = b^2\,\mathrm{Var}[x],$$

which implies, for any constant a, that

$$\mathrm{Var}[a] = 0.$$

SOME SPECIFIC PROBABILITY DISTRIBUTIONS

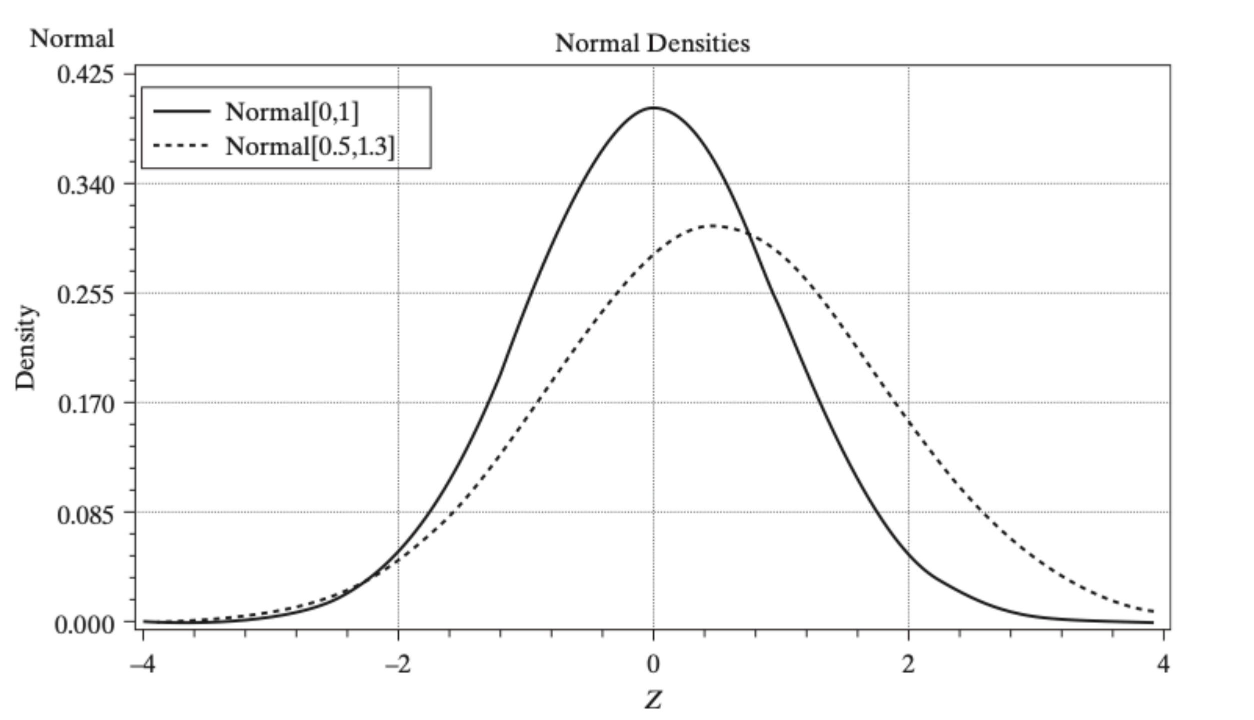
The general form of the normal distribution with mean μ and standard deviation σ is

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

This result is usually denoted $x \sim N[\mu, \sigma^2]$.

standard normal distribution, denoted $N[0, 1]$.

THE NORMAL AND SKEW NORMAL DISTRIBUTIONS



SOME SPECIFIC PROBABILITY DISTRIBUTIONS

If $z \sim N[0, 1]$, then $x = z^2 \sim$ chi-squared[1]—that is, **chi-squared** with one degree of freedom—denoted

$$z^2 \sim \chi^2[1].$$

If $x_1, \ldots, x_n$ are n *independent* chi-squared[1] variables, then

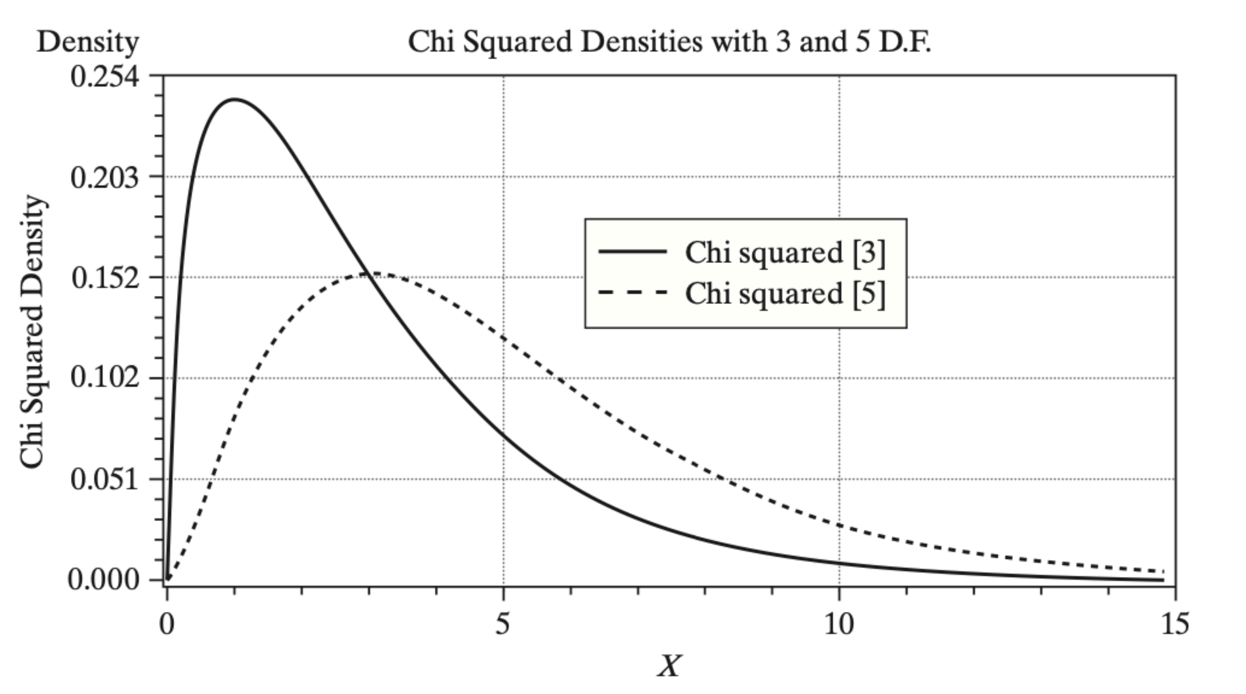
$$\sum_{i=1}^{n} x_i \sim \text{chi-squared}[n].$$

If x_1 and x_2 are two *independent* chi-squared variables with degrees of freedom parameters x_1 and x_1 respectively, then the ratio

$$F[n_1, n_2] = \frac{x_1/n_1}{x_2/n_2}$$

has the ***F* distribution** with n_1 and n_2 degrees of freedom.

CHI-SQUARED, *t*, AND *F* DISTRIBUTIONS



SOME SPECIFIC PROBABILITY DISTRIBUTIONS

If z is an $N[0,1]$ variable and x is $\chi^2[n]$ and is independent of z, then the ratio

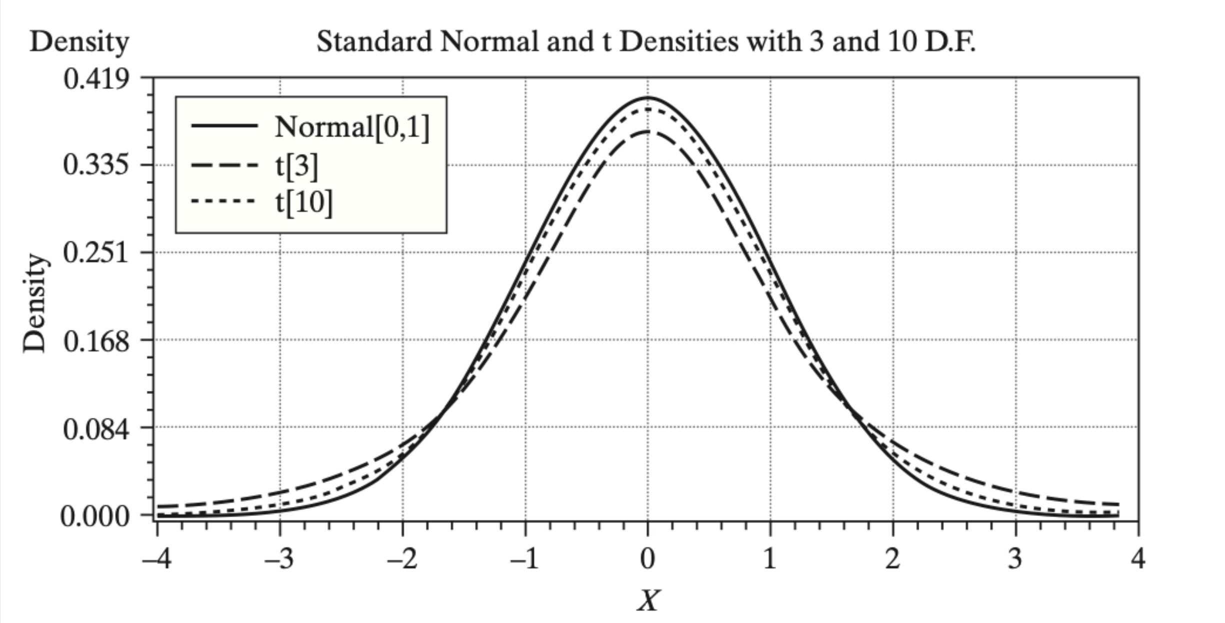
$$t[n] = \frac{z}{\sqrt{x/n}}$$

has the t **distribution** with n degrees of freedom.

The t distribution has the same shape as the normal distribution but has thicker tails.

If $t \sim t[n]$, then $t^2 \sim F[1,n]$.

CHI-SQUARED, *t*, AND *F* DISTRIBUTIONS



Standardization of Normal Distribution

One of the most useful applications of the preceding result is the linear transformation of a normally distributed variable. If $x \sim N[\mu, \sigma^2]$, then the distribution of

$$y = \frac{x - \mu}{\sigma}$$

JOINT DISTRIBUTIONS

The **joint density function** for two random variables X and Y denoted $f(x,y)$ is defined so that

$$\text{Prob}(a \le x \le b, c \le y \le d) = \begin{cases} \sum_{a \le x \le b} \sum_{c \le y \le d} f(x, y) & \text{if } x \text{ and } y \text{ are discrete,} \\ \int_a^b \int_c^d f(x, y)\, dy\, dx & \text{if } x \text{ and } y \text{ are continuous.} \end{cases}$$

MARGINAL DISTRIBUTIONS

A **marginal probability density** or marginal probability distribution is defined with respect to an individual variable. To obtain the marginal distributions from the joint density, it is necessary to sum or integrate out the other variable:

$$f_x(x) = \begin{cases} \sum_y f(x, y) & \text{in the discrete case} \\ \int_y f(x, s)\, ds & \text{in the continuous case,} \end{cases}$$

and similarly for $f_y(y)$.

MARGINAL DISTRIBUTIONS

Two random variables are statistically independent if and only if their joint density is the product of the marginal densities:

$$f(x, y) = f_x(x) f_y(y) \Leftrightarrow x \text{ and } y \text{ are independent.}$$

EXPECTATIONS IN A JOINT DISTRIBUTION

The means, variances, and higher moments of the variables in a joint distribution are defined with respect to the marginal distributions. For the mean of x in a discrete distribution,

$$
\begin{aligned}
E[x] &= \sum_x x f_x(x) \\
&= \sum_x x \left[\sum_y f(x, y) \right] \\
&= \sum_x \sum_y x f(x, y).
\end{aligned}
$$

EXPECTATIONS IN A JOINT DISTRIBUTION

The means of the variables in a continuous distribution are defined likewise, using integration instead of summation:

$$E[x] = \int_x x f_x(x)\,dx$$

$$= \int_x \int_y x f(x, y)\,dy\,dx.$$

Variances are computed in the same manner:

$$\operatorname{Var}[x] = \sum_x (x - E[x])^2 f_x(x)$$

$$= \sum_x \sum_y (x - E[x])^2 f(x, y).$$

COVARIANCE AND CORRELATION

For any function $g(x, y)$,

$$E[g(x, y)] = \begin{cases} \sum_x \sum_y g(x, y) f(x, y) & \text{in the discrete case} \\ \int_x \int_y g(x, y) f(x, y)\, dy\, dx & \text{in the continuous case.} \end{cases}$$

The covariance of x and y is a special case:

$$\begin{aligned} \operatorname{Cov}[x, y] &= E[(x - \mu_x),(y - \mu_y)] \\ &= E[xy] - \mu_x \mu_y \\ &= \sigma_{xy}. \end{aligned}$$

COVARIANCE AND CORRELATION

If x and y are independent, then $f(x, y) = f_x(x)f_y(y)$ and

$$\begin{aligned}
\sigma_{xy} &= \sum_x \sum_y f_x(x)f_y(y)(x - \mu_x)(y - \mu_y) \\
&= \sum_x (x - \mu_x)f_x(x) \sum_y (y - \mu_y)f_y(y) \\
&= E[x - \mu_x]E[y - \mu_y] \\
&= 0.
\end{aligned}$$

COVARIANCE AND CORRELATION

If x and y are independent, then $f(x, y) = f_x(x)f_y(y)$ and

$$\begin{aligned}
\sigma_{xy} &= \sum_x \sum_y f_x(x)f_y(y)(x - \mu_x)(y - \mu_y) \\
&= \sum_x (x - \mu_x)f_x(x) \sum_y (y - \mu_y)f_y(y) \\
&= E[x - \mu_x]E[y - \mu_y] \\
&= 0.
\end{aligned}$$

COVARIANCE AND CORRELATION

The sign of the covariance will indicate the direction of covariation of X and Y. Its magnitude depends on the scales of measurement, however. In view of this fact, a preferable measure is the correlation coefficient:

$$r[x, y] = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y},$$

where σ_x and σ_y are the standard deviations of x and y, respectively. The correlation coefficient has the same sign as the covariance but is always between -1 and 1 and is thus unaffected by any scaling of the variables.