

Appendix B — Some Useful Probability Distributions

B.1 Introduction

The following is a short overview of some of the most common probability distributions encountered in the context of linear regression.

B.2 Normal distribution

The normal or Gaussian distribution is of principal importance in probability and statistics. In its univariate form, we say that $X \sim \mathcal{N}(\mu, \sigma^2)$ with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$ if it has the following probability density function (pdf):

$$f_{\mathcal{N}(\mu, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right).$$

Such a random variable X can be centred and scaled into a standardized form as $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$. Furthermore, the normal distribution is very stable in the sense that the sum of a collection of independent normal random variables has a normal distribution as well as scaling a normal random variable by some real valued scalar.

Part of its importance stems from the central limit theorem, which in its simplest form states that the standardized sum of a collection on independent random variables converges in distribution to a standard normal random variable. However, its worth noting that there are many other central limit theorems in existence.

Theorem B.1 (Central Limit Theorem) Let Y_1, \dots, Y_n be a sample of n iid random variables with mean $\mu \in \mathbb{R}$ and finite variance $\sigma^2 < \infty$. Letting $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ be the sample mean, then

$$\sqrt{n}(\bar{Y} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where \xrightarrow{d} denotes convergence in distribution.

The normal distribution can be extended to the multivariate normal distribution for a vector $X \in \mathbb{R}^p$ where we write $X \sim \mathcal{N}(\mu, \Sigma)$. Here, $\mu \in \mathbb{R}^p$ is the mean vector while $\Sigma \in \mathbb{R}^{p \times p}$ is a $p \times p$ matrix that is symmetric and positive semi-definite. This distribution also has elliptical symmetry. The form of the pdf, assuming Σ is positive definite, is

$$f_{\mathcal{N}(\mu, \Sigma)}(x) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

Otherwise, X will have a degenerate normal distribution, which is still normal but supported on a subspace of dimension less than p .

The multivariate normal (MVN) distribution has some very nice characterizations. A vector $X = (X_1, \dots, X_p)$ is MVN if and only if every linear combination of the X_i is univariate normal. That is, for all $a \in \mathbb{R}^p$,

$a \cdot X \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$. Hence, it is worth emphasizing that if X is a vector that is marginally normal (i.e. every component X_i is univariate normal), then the joint distribution is **not** necessarily multivariate normal.

In general if two random variables $X, Y \in \mathbb{R}$ are independent, then we have that $\text{cov}(X, Y) = 0$. However, the reverse implication is not necessarily true. One exception to this is that if (X, Y) is MVN then if $\text{cov}(X, Y) = 0$ then X and Y are independent.

B.3 Chi-Squared distribution

The chi-squared distribution arises throughout the field of statistics usually in the context of goodness-of-fit testing. Let $Z \sim \mathcal{N}(0, 1)$, then $Z \sim \chi^2(1)$, which is said to be chi-squared with one degree of freedom. Furthermore, chi-squared random variables are additive in the sense that if $X \sim \chi^2(\nu)$ and $Y \sim \chi^2(\eta)$ are independent, then $X + Y \sim \chi^2(\nu + \eta)$.

A chi-squared random variable is supported on the positive real line and the pdf for $X \sim \chi^2(\nu)$ with $\nu > 0$ is

$$f_{\chi^2(\nu)}(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2}$$

Its mean and variance are ν and 2ν , respectively. Note that while very often the degrees of freedom parameter is a positive integer, it can take on any positive real value and still be valid. In fact, the chi-squared distribution is a specific type of the more general gamma distribution, which will not be discussed in these notes.

This random variable is quite useful in the context of linear regression with respect to how it leads to the t and F distributions discussed in the following subsections. However, it also arises in many other areas of statistics including Wilks' Theorem regarding the asymptotic distribution of the log likelihood ratio, goodness-of-fit in multinomial hypothesis testing, and testing for independence in a contingency table.

B.4 t distribution

The t distribution is sometimes referred to as Student's t distribution in recognition of its, at the time, anonymous progenitor William Sealy Gosset working at the Guinness brewery. It most notably arises in the context of estimation of a normal random variable when the variance is unknown as occurs frequently in these lecture notes.

Let $Z \sim \mathcal{N}(0, 1)$ and let $V \sim \chi^2(\nu)$ be independent random variables. Then, we say that

$$T = \frac{Z}{\sqrt{V/\nu}} \sim t(\nu)$$

has a t distribution with ν degrees of freedom. Such a distribution can be thought of as a heavier tailed version of the standard normal distribution. In fact, $t(1)$ is the Cauchy distribution with pdf

$$f_{t(1)}(x) = (\pi(1+x^2))^{-1}$$

while $T \sim t(\nu)$ converges to a standard normal distribution as $\nu \rightarrow \infty$.

A noteworthy property of a t distributed random variable is that it will only have moments up to but not including order ν .

That is, for $T \sim t(\nu)$, $\text{E}T^k < \infty$ for $k < \nu$.

For $k \geq \nu$, the moments do not exist.

B.5 F distribution

The F distribution arises often in the context of linear regression when a comparison is made between two sources of variation. Let $X \sim \chi^2(\nu)$ and $Y \sim \chi^2(\eta)$ be independent random variables, then we write that

$$F = \frac{X/\nu}{Y/\eta} \sim F(\nu, \eta)$$

has an F distribution with degrees of freedom ν and η . Due to this form of the distribution, if $F \sim F(\nu, \eta)$, then $F^{-1} \sim F(\eta, \nu)$. The F distribution is supported on the positive real line. The F and t distributions are related by the fact that if $T \sim t(\nu)$, then $T^2 \sim F(1, \nu)$.