Chapter 7

Sampling Distributions

In Chapter 2, we discussed how statistics are random and parameters are constant. We now combine our prior notions of statistics with random variables to form our theory about inferential statistics. Instead of thinking about statistics as a random number, we can think about these numbers realizations of a random variable and the statistic itself as a random variable. For example, the sample mean is one statistic we considered. It is a function of the realizations of some underlying random variable, which represents the data. We take these realizations of the underlying random variable and the sample mean statistic adds them and divides by the number of realizations. We build the sample mean off of a random variable, so it is also random. Like all random variables, statistics possess probability distributions. We call the probability distribution for a statistic its *sampling distribution.* The sampling distribution is the probability distribution defined over the support of a statistic. The support of most statistics we consider is the set of real numbers or the set of positive real numbers. The major task of inferential statistics is to find the sampling distribution for a statistic from which we can make probabilistic statements (i.e. inference).

7.1 *i.i.d.* Random Samples

Definition Let $x_1, x_2, x_3, ..., x_n$ be realizations of the random variables $X_1, X_2, X_3, ..., X_n$, respectively. We say $x_1, x_2, x_3, ..., x_n$ is a random sample. If the random variables $X_1, X_2, X_3, ..., X_n$ are independent, then we say $X_1, X_2, X_3, ..., X_n$ are independently distributed. If the random variables $X_1, X_2, X_3, ..., X_n$ have the same underlying distribution, then we say $X_1, X_2, X_3, ..., X_n$ are identically distributed. We call independently and identically distributed random variables *i.i.d.* for short.

Suppose $f(x_1, x_2, x_3, ..., x_n)$ is the joint probability density function for the continuous random variables $X_1, X_2, X_3, ..., X_n$. When $X_1, X_2, X_3, ..., X_n$ is independently distributed, we know the joint density function is the product of the marginal densities

$$f(x_1, x_2, x_3, ..., x_n) = \Pi_{i=1}^n f_i(x_i).$$

We subscript f by the random variable i because the distributions underlying the random variables could be different. In the case that $X_1, ..., X_n$ is identically distributed, then $f_i = f_j \; \forall i, j$. Hence, the joint PDF for *i.i.d.* random variables $X_1, ..., X_n$ is

$$f(x_1, x_2, x_3, ..., x_n) = \Pi_{i=1}^n f(x_i).$$

Example Let X_1 and X_2 be independent and identically distributed random variables with mean μ and variance σ^2. Then probability density function defined over all possible realizations of (x_1, x_2) is

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= f_X(x_1) \cdot f_X(x_2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x_1-\mu}{\sigma})^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x_2-\mu}{\sigma})^2} \\ &= \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\sum\limits_{i=1}^{2}(\frac{x_i-\mu}{\sigma})^2} \end{aligned}$$

7.2 Sample Mean and Variance

Recall two statistics we defined earlier. Now instead of data, we define these statistics over random variables. The actual realization of the statistic will depend on the realizations of the underlying random variables.

Definition Let $X_1, ..., X_n$ be random variables. Then the sample mean is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Note that the sample mean and sample variance are random variables, because we construct them from random variables $X_1, ..., X_n$. When we have a sample of data, we observe one realization of these random variables. When $X_1, ..., X_n$ are *i.i.d.*, the standard moments (mean and variance) of the sample mean have nice expressions.

Proposition 7.2.1 *Suppose X is independently and identically distributed for all i so that $\mathbb{E}[X_i] = \mu$ and $Var[X_i] = \sigma^2$. Then $\mathbb{E}[\bar{X}] = \mu$ and $Var[\bar{X}] = \frac{\sigma^2}{n}$.*

Proof

$$\begin{aligned}
\mathbb{E}[\bar{X}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}[X_i]}_{=\mu} \\
&= \frac{1}{n}\cdot \mu \sum_{i=1}^{n} 1 \\
&= \frac{1}{n}\mu \cdot n \\
&= \mu
\end{aligned}$$

$$
\begin{aligned}
Var[\bar{X}] &= Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\cdot Var\left[\sum_{i=1}^{n} X_i\right] \\
\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] &= n\mu \\
\Rightarrow Var[\bar{X}] &= \frac{1}{n^2}\cdot \mathbb{E}\left[(X_1 + X_2 + ... + X_n - n\mu)^2\right] \\
&= \frac{1}{n^2}\cdot \mathbb{E}\left[((X_1-\mu) + (X_2-\mu) + ... + (X_n-\mu))^2\right] \\
&= \frac{1}{n^2}\cdot \mathbb{E}\left[(X_1-\mu)^2 + (X_1-\mu)(X_2-\mu) + ... + (X_n-\mu)^2\right] \\
&= \frac{1}{n^2}\cdot \left[\sum_{i=1}^{n}\mathbb{E}(X_i-\mu)^2 + 2\sum_{i}\sum_{j\neq i}(X_1-\mu)(X_2-\mu)\right] \\
&= \frac{1}{n^2}\cdot \left[\sum_{i=1}^{n} Var(X_i) + 2\sum_{i}\sum_{j\neq i} Cov(X_i, X_j)\right] \\
Cov(X_i, X_j) &= 0 \;\forall i \neq j \text{ (because } X_i \text{ is } i.i.d.\text{)} \\
\Rightarrow &= \frac{1}{n^2}\cdot\left[\sum_{i=1}^{n}\sigma^2\right] \\
\Rightarrow Var[\bar{X}] &= \frac{\sigma^2}{n} \quad \blacksquare
\end{aligned}
$$

Remark For the sample mean of *i.i.d.* random variables, $Var[\bar{X}] \to 0$ as $n \to \infty$. In other words, the variance in the sample mean collapses to zero as the sample size grows. Where does it collapse? It collapses onto the mean of the $\bar{X}$ which also happens to be the mean of X_i ($\mathbb{E}[X_i] = \mu$). The sample mean collapses probabilistically onto the true mean of the underlying random variable (the random variables are *i.i.d.*, so there is really only one underlying random variable taking on n realizations). We call this property *mean-squared error consistency*, which we will discuss more later. The sample mean also has another special property when constructed from an *i.i.d.* random variable. Its expected value is the true population mean for the underlying random variable. This property means that over repeated samples of size n, the sample mean will be the true population mean on average. We call this property *unbiasedness*, which we will also discuss more later.

7.3 Normal Random Variables

Recall the PDF for a normally distributed random variable x with expected value μ and variance σ^2 is

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

Looking at the PDF (and thereby kernel of the CDF) shows that the mean and the variance characterise the entire distribution for a normal random variable. A distribution is generally characterised by every moment, however the first moment (the mean) and the standardised second moment (the variance) are sufficient parameters to describe normal random variables. This property is one of several special properties of the normal distribution.

Generally, we have the following relationship chain for random variables

$$\text{Independence} \Rightarrow \text{Mean Independence} \Rightarrow \text{Uncorrelated},$$

however with *uncorrelated, jointly normally distributed* random variables the relationship goes in every direction

$$\text{Independence} \Leftrightarrow \text{Mean Independence} \Leftrightarrow \text{Uncorrelated.}$$

The relationship goes in each direction due to special form of the joint normal PDF which factors into the product of marginals as long as the correlation between the random variables of the joint PDF is zero.

Definition We say the random variables $X_1, X_2, ..., X_n$ are mean independent when

$$\mathbb{E}[X_1 \cdot X_2 \cdot ... \cdot X_n] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2] \cdot ... \cdot \mathbb{E}[X_n].$$

Definition The random vector consisting of $X_1, X_2, ..., X_n$ is normally distributed when the random variable components are jointly normally distributed where

$$f(x_1, ..., x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^{-n} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

where x is the vector $(x_1, ..., x_n)^T$ and μ is the vector $(\mu_1, ..., \mu_n)$, corresponding to the expected value for each random variable component. Σ is the so-called variance-covariance matrix of the random vector, which is defined by $\mathbb{E}[(x-\mu)(x-\mu)^T]$ (the multivariate analogue of the variance). It is symmetric and positive definite.

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & . & . & . & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & . & . & . & Cov(X_2, X_n) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ Cov(X_n, X_1) & Cov(X_n, X_2) & . & . & . & Var(X_n) \end{pmatrix}$$

Uncorrelated random variables have zero-covariance. From the PDF of jointly normally distributed random variables, zero-covariances mean Σ is a diagonal matrix. We can factor the joint into the product of marginals for each X_i when Σ is diagonal. Hence, uncorrelated implies independence and vice versa in this special case.

7.4 The Central Limit Theorem

We can often calculate the sample analogues of these standardised moments with ease, and they often are quite useful. We only know the entire distribution of a random variable from its mean and variance when the random variable is normal. However, the sample mean with *i.i.d.* random variables exhibits a special characteristic. If we normalise the sample mean properly, then we obtain a distribution for the sample mean statistic that is approximately normal. This property comes from a result we call the *central limit theorem.* The central limit theorem tells us that the sample mean (which is a random variable!) when normalised properly converges to a standard normal distribution as the sample size grows, regardless of the underlying distribution of the random variable from which we construct the sample mean. This result is astounding! We can take a random variable with any distribution, make *i.i.d.* draws from this variable and take the sample mean. As the sample size grows, the distribution of the *sample mean* will be approximately a standard normal distribution characterised solely by the mean and variance of the sample mean, which we just calculated!

Proposition 7.4.1 (Central Limit Theorem for *i.i.d.* Random Variables) *Suppose X_i is distributed i.i.d. for all i. Let $\bar{X}$ be the sample mean and (μ, σ^2) be the mean and variance of X, respectively. Then*

$$\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \to_d N(0,1).$$

Hence, $\sqrt{n}\bar{X} \to_d N(\mu, \sigma^2)$.

Let's illustrate the CLT to get a sense of how it works. Suppose X is a discrete random variable where

$$X = \begin{cases} 1 & w.p.\ 1/3, \\ 2 & w.p.\ 1/3, \\ 3 & w.p.\ 1/3. \end{cases}$$

Then we can illustrate the PMF of X as follows:

$$\begin{array}{|ccc} \ast & \ast & \ast \\ \hline 1 & 2 & 3 \end{array}$$

Now we can take the average of two independent draws from X to obtain the sample mean PMF

$$p(\bar{X}) = \begin{cases} 1/9, & \bar{X} = 1 \ or \ \bar{X} = 3 \\ 2/9, & \bar{X} = 1.5 \ or \ \bar{X} = 2.5 \\ 3/9, & \bar{X} = 2 \end{cases}$$

which now looks like

$$\begin{array}{|ccccc} & & \ast & & \\ & \ast & \ast & \ast & \\ \ast & \ast & \ast & \ast & \ast \\ \hline 1 & 1.5 & 2 & 2.5 & 3 \end{array}$$

Averaging of three independent draws the sample mean PMF is

$$p(\bar{X}) = \begin{cases} 1/27, & \bar{X} = 1 \ or \ \bar{X} = 3 \\ 3/27, & \bar{X} = 1.\bar{3} \ or \ \bar{X} = 2.\bar{6} \\ 6/27, & \bar{X} = 1.\bar{6} \ or \ \bar{X} = 2.\bar{3} \\ 7/27, & \bar{X} = 2 \end{cases}$$

which now looks like

$$\begin{array}{|ccccccc} & & & \ast & & & \\ & & \ast & \ast & \ast & & \\ & & \ast & \ast & \ast & & \\ & & \ast & \ast & \ast & & \\ & \ast & \ast & \ast & \ast & \ast & \\ & \ast & \ast & \ast & \ast & \ast & \\ \ast & \ast & \ast & \ast & \ast & \ast & \ast \\ \hline 1 & 1.\bar{3} & 1.\bar{6} & 2 & 2.\bar{3} & 2.\bar{6} & 3 \end{array}$$

We only increase n to three and already have something that starts to looked shaped like a normal distribution centered on the true mean of the random variable X. We can even calculate $Var(\bar{X})$ using the PMF to see that it is $\sigma^2/3$.

7.5 Derivatives of the Normal Distributions

Suppose we take the sum or difference of two independent normal random variables. We obtain a new random variable which is also normally distributed with a mean equal to the sum of the means of the normal random variables and a variance equal to the sum of their variances. We can verify this fact using the multivariate transformation H where $(U, V) = H(X, Y), U = X + Y, V = Y, X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2)$ and X and Y are independent to obtain $f_U(u)$. Normal random variables have many special transformations like the one above when the mean is zero and the variance is one (i.e. a standard normal random variable). These new distributions resulting from transformations of a normal random variable are part of a collection of distributions associated with normal distribution. We often appeal to the central limit theorem for a statistic to be approximately normal, then we look at distributions within this family to make inference. These distributions will be asymptotic approximations to the exact finite sample distribution when we apply the central limit theorem and exact sampling distributions when the random variable is normally distributed.

7.5.1 Chi-Squared Distribution

Definition Suppose Z_i are k independent standard normal random variables for every i. Then

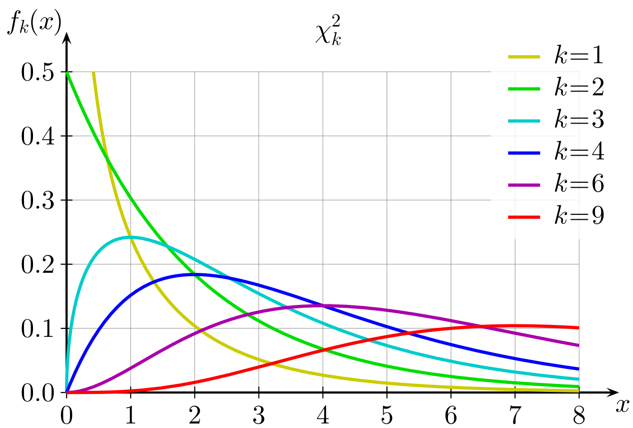
$$\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$$

that is the sum of the squares of k independent standard normal random variables is a *chi-squared* distributed random variable with k degrees of freedom. The PDF for a chi-squared random variable is

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

where $\Gamma(\cdot)$ is the *gamma function*. The chi-squared random variable with degrees of freedom k has mean k and variance $2k$ with support $[0, \infty)$.

The number of degrees of freedom has various definitions. We will think of it as a parameter describing a distribution. In our definitions, the degrees of freedom are the number of independent pieces of information entering a statistic less the number of parameters estimated in constructing the statistic. This definition is precise but not formal.



We will often use the chi-squared distribution to make inference about the variance of a random variable from the sample variance.

Example Suppose X_i is *i.i.d.* $N(\mu, \sigma^2)$, then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. Why?

We can show $Cov(\bar{X}, X_i - \bar{X}) = 0$ for all i, which implies $\bar{X}$ and S^2 are independent because (1) $(\bar{X}, X_1 - \bar{X}, ..., X_2 - \bar{X})$ are jointly normally distributed and (2) if X and Y are independent, then any function of X is also independent of Y (S^2 is a function of $X_i - \bar{X}$).

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\
\Rightarrow \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 &= \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2,
\end{aligned}$$

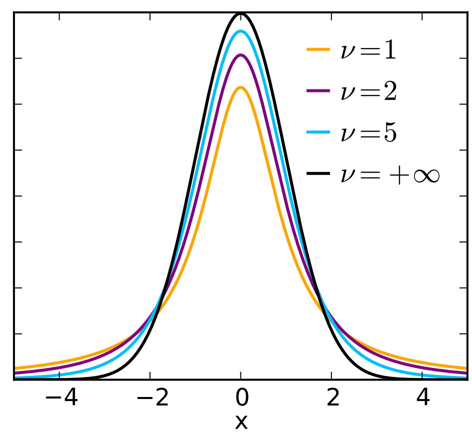
dividing by σ^2 on both sides and substituting S^2 for $\sum_{i=1}^{n}(X_i - \bar{X})^2$. Now the right hand side is chi-squared random variables with n degrees of freedom, because $\frac{X_i - \mu}{\sigma}$ is a standard normal random variable. Similarly, $\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ is a chi-squared random variable with 1 degree of freedom. From here, we can deduce that the left hand side is the sum of two independent chi-squared random variables - one with 1 degree of freedom and hence the other must have $n-1$ degrees of freedom.

7.5.2 Student's-t Distribution

Definition Suppose Z are Y independent random variables where $Z \sim N(0,1)$ and $Y \sim \chi^2(\nu)$ Then

$$\frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

that is the ratio a standard normal random variable and the square root of independent chi-squared random variable divided by its degrees of freedom (ν) forms a Student's-t distribution with ν degrees of freedom.



As the degrees of freedom increases, the Student's-t distribution converges to the standard normal distribution.

Example Suppose X_i is $i.i.d.$ $N(\mu, \sigma^2)$, then $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$. Why? Rearranging,

$$\begin{aligned}
\frac{\sqrt{n}(\bar{X}-\mu)}{S} &= \frac{\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} \\
&= \frac{\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}} \\
&= \frac{\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1}\frac{(n-1)S^2}{\sigma^2}}} \\
&= \frac{\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1}\frac{(n-1)S^2}{\sigma^2}}} \\
Z &\equiv \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0,1) \\
Y &\equiv \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \\
\Rightarrow &= \frac{Z}{\sqrt{Y/(n-1)}} \sim t_{n-1}
\end{aligned}$$

7.5.3 F-Distribution

Definition Suppose W are Y independent random variables where $W \sim \chi^2(\lambda)$ and $Y \sim \chi^2(\nu)$. Then

$$\frac{W/\lambda}{Y/\nu} \sim F_{\lambda,\nu}$$

that is the ratio of two independent chi-squared distribution divided by their degrees of freedom is a F-distributed random variable.

We often use the F-distribution to make inference on the equality of two variances of two independent random variables and to test whether certain restriction hold. We will say more about the F-distribution later.

Example Suppose S^1 and S^2 are the sample variances of independent random sample of size n_1 and n_2 for normal populations (or random variables) with variances σ_1^2 and σ_2^2. Then $\sigma_2^2 S_1^2/\sigma_1^2 S_2^2 \sim F_{n_1-1,n_2-1}$. Why?

$$\begin{aligned}
(n_1-1)S_1^2/\sigma_1^2 &\sim \chi^2(n_1-1) \\
(n_2-1)S_2^2/\sigma_2^2 &\sim \chi^2(n_2-1) \\
\Rightarrow \frac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2} &\sim \frac{\chi^2(n_1-1)}{\chi^2(n_2-1)} \\
\Rightarrow \frac{n_2-1}{n_1-1}\cdot\frac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2} &\sim \frac{n_2-1}{n_1-1}\cdot\frac{\chi^2(n_1-1)}{\chi^2(n_2-1)} \\
\Leftrightarrow \frac{n_2-1}{n_1-1}\cdot\frac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2} &\sim \frac{\chi^2(n_1-1)/(n_1-1)}{\chi^2(n_2-1)/(n_2-1)} \\
&= F_{n_1-1,n_2-1} \\
\Rightarrow \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} &\sim F_{n_1-1,n_2-1}
\end{aligned}$$