6 Distributions Derived from the Normal Distribution

Definition. *The random variable $X_1, \cdots, X_n$ are called a* ***random sample of size*** n ***form the population*** *if $X_1, \cdots, X_n$ are all independent random variables and the marginal pdf (continuous case) or pmf (discrete case) of each X_i is $f(x)$. In other words,*

$$X_1, \cdots, X_n \overset{iid}{\sim} f(x)$$

Definition. *Any function $g(X_1, \cdots, X_n)$ of the sample is called a (sample) statistics.*

Definition. *The distribution of a statistic $T = g(X_1, \cdots, X_n)$ is called the* ***sampling distribution*** *of T*

Definition. *Let $X_1, \cdots, X_n$ be a random sample from some distribution/population. Then* ***sample mean*** *and* ***sample variance*** *is*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

The sampling distribution can be partially characterized by

$$E[\overline{X}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} \cdot n\mu = \mu$$

$$Var(\overline{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Note that in practice, we can compute

Remark. If $X_i \sim \mathbb{N}(\mu, \sigma^2)$ then

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

If n is large, even if X_i are not normal, by CLT we have

$$\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \iff P(\overline{X} \leq t) \approx \Phi(\frac{t-\mu}{\sigma/\sqrt{n}})$$

Definition. ***chi-squred distribution*** *If $Z \sim \mathcal{N}(0,1)$, the distribution of $U = Z^2$ is called the chi-squared distribution with 1 degree of freedom. Let $Z_1, \cdots, Z_n \overset{iid}{\sim} \mathcal{N}(0,1)$, the distribution of statistic $X^2 = \sum_{i=1}^{n} Z_i^2$ is called* ***chi-squred distribution with*** n ***degrees of freedom*** *and is denoted by χ_n^2*

Remark.

1. $\chi_1^2 = \Gamma(\frac{1}{2}, \frac{1}{2})$ and $\chi_n^2 = \Gamma(\frac{n}{2}, \frac{1}{2})$ and therefore $E[\chi_n^2] = n$ and $Var[\chi_n^2] = 2n$

2. MGF of $Y \sim \chi_n^2$ is $M_Y(t) = (1-2t)^{-n/2}$ (note mgf for gamma $(1-\theta t)^{-k}$)

3. $X \sim \chi_m^2$ and $Y \sim \chi_n^2$ then $X + Y \sim \chi_{m+n}^2$

Definition. ***t distribution*** *If $Z \sim \mathcal{N}(0,1)$ and $U \sim \chi_n^2$ and Z and U are independent, then the distribution of $Z/\sqrt{U/n}$ is called the t distribution with n degrees of freedom.*

Remark. 1. density function is symmetric, i.e. $f(x) = f(-x)$.

2. As $n \to \infty$, $Z \overset{d}{\sim} \mathcal{N}(0,1)$

Definition. ***F distribution*** *Let U and V be independent chi-squared random variable with m and n degrees of freedom. The distribution of*

$$W = \frac{U/m}{V/n}$$

is called F distribution with m and n degrees of freedom and denoted by $F_{m,n}$

Remark. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be two independent random samples from a $\mathcal{N}(\mu, \sigma^2)$ distribution and let S_X^2 and S_Y^2 be their sample variances. Then

$$F = \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

Theorem. ***independence of $\overline{X}$ and S^2*** *Let $X_1, \cdots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ and let $\overline{X}$ and S^2 be sample mean and variance, then $\overline{X}$ and S^2 are independent*

Theorem. ***relationship between S^2 and chi-squared distribution*** *Let $X_1, \cdots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ and let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ be sample variance, then*

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Corollary. ***sample standardization converges to t distribution*** *Let $\overline{X}$ and S^2 be given, then*

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Proof.

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} = \frac{\mathcal{N}(0,1)}{\sqrt{\chi_{n-1}^2/(n-1)}} \sim t_{n-1}$$

□

8 Parameter Estimation and Fitting of Probability Distribution

Definition. ***estimator of parameter*** *Let $X_1, \cdots, X_n$ be a random sample from some distribution, and let θ be a parameter of that distribution. Any statistic $U = U(X_1, \cdots, X_n)$ that is used to estimate θ is called an estimator of θ, $\hat{\theta}$*

Definition. ***Method of Moments Estimator*** *The kth moment of a distribution is defined as*

$$\mu_k = E[X^k]$$

The kth ***sample moment*** *is defined as*

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Suppose the goal is to estimate $\theta = (\theta_1, \cdots, \theta_p)$ characterizing the distribution $f_X(x|\theta)$. The methods of moment estimator of is the solution to

$$\mu_i(\hat{\theta}) = m_i \ i = 0, \cdots, p$$

Remark. The idea is by using this system of equation, we express population parameters θ with parameters. Then replace known parameter by estimates $\hat{\theta}$

The following comes up in simplifying equality for second moments $m_2 - m_1 = \sigma^2$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}^2 + \overline{X}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{2\overline{X}}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n}\overline{X}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 && \text{(looks similar to } S^2\text{)}
\end{aligned}$$

Definition. ***Consistent Estimator*** *Let $X_1, \cdots, X_n \sim f_\theta$ (a sample from a distribution characterized by θ). We say that $\hat{\theta_n} = \hat{\theta_n}(X_1, \cdots, X_n)$ is a consistent estimator of θ if $\hat{\theta_n}$ converges in probability to θ as n approaches infinity; that is, for any $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat{\theta_n} - \theta\right| > \epsilon\right) = 0$$

Remark. Remember how weak law of large number (LLN) implies that sample moments converges in probability to population moments, that is for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| > \epsilon\right) = 0$$

Or equivalently, sample moment $m_1 = \overline{X}$ is a consistent estimator for population moment $\mu_1 = \mathbb{E}[X]$. Since function relating estimates to sample moments are continuous, the estimates will converge to parameters as the sample moments converge to population moments. In general for any k

$$m_k \xrightarrow{p} \mu_k$$

Theorem. ***Property for Convergence in Probability*** *Let $\hat{\theta}_n \xrightarrow{p} \theta$ and $\hat{\eta}_n \xrightarrow{p} \eta$*

1. $\hat{\theta}_n + \hat{\eta}_n \xrightarrow{p} \theta + \eta$
2. $\hat{\theta}_n \hat{\eta}_n \xrightarrow{p} \theta\eta$
3. $g(\hat{\theta}_n) \xrightarrow{p} g(\theta)$ *for any continuous* g

Remark. Method of Moments Estimators are usually consistent. Take MME estimator for normal distribution.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = m_2 - m_1^2$$

By Weak Law of Large Numbers, $m_1 = \overline{X} \xrightarrow{p} \mathbb{E}[X]$ and by its generalization $m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} \mathbb{E}[X^2]$, so by property for convergence in probability

$$\hat{\sigma}^2 = m_2 - m_1^2 \xrightarrow{p} \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sigma^2$$

implies that both $\hat{\mu}$ and $\hat{\sigma^2}$ are consistent estimator for normal distribution. Note that sample variance is also consistent

$$S^2 = \frac{n}{n-1}\hat{\sigma^2} \xrightarrow{p} 1 \cdot \sigma^2 = \sigma^2$$

The Method of Maximum Likelihood

Definition. ***Likelihood Function*** *Let $X_1, \cdots, X_n$ be continuous random variables with joint pdf $f(x_1, \cdots, x_n|\theta)$, where θ is a parameter. For a given vector of observations $(x_1, \cdots, x_n)$, the probability of observing the given data as a function of parameter θ, i.e. the likelihood function, is given by*

$$\mathcal{L} = f(x_1, \cdots, x_n|\theta)$$

If X_1 are assumed to be i.i.d., their joint density is the product of marginal densities

$$\mathcal{L} = \prod_{i=1}^{n} f(X_i = x_i|\theta)$$

For convinience of maximizing $\mathcal{L}$ we maximize the log likelihood

$$l(\theta) = \sum_{i=1}^{n} \log f(X_i = x_i|\theta)$$

Definition. *The* ***Maximum Likelihood Estimator (MLE)*** *of* θ *is*

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \mathcal{L}(\theta)$$

Remark. We can also maximize log likelihood $l(\theta)$. For simplificity sake, we compute θ as critical points which satisfies the condition that first order partial $l'(\theta) = 0$ in order to find the maximum. We then use second derivative test to verify it is indeed the maximum point, i.e. $l^{''}(\theta) < 0$ One thing to note that the final $\hat{\theta}$ should not depend on population parameter θ but be expressed entirely of sample statistics

Definition. *The* ***Newton-Raphson Method*** *finds successively better approximations to the roots (or zeroes) of a real-valued function* $x : f(x) = 0$. *The following process is repeated until a sufficiently accurate value* x_{n+1} *is reached*

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

where x_0 *is the initial guess for the root of function*

Remark. Sometimes there is no closed form solution (i.e. maximum value) to $\dfrac{\partial l}{\partial \theta} = 0$. We use Newton-Raphson method to find $\hat{\theta}$ that satisfies the function. We iterate over θ to convergence

$$\hat{\theta}_{new} = \hat{\theta}_{old} - \frac{l'(\hat{\theta}_{old})}{l''(\hat{\theta}_{old})}$$