STA257 Probability and Statistics I

Gun Ho Jang

Update on June 18, 2018

Note: This note is prepared for STA257. There might be numerous fault arguments/statements/typos. If you spot one, please contact the instructor or you may look up references which may contain errors too.

Normal and related distributions

Normal distributions are one of the most commonly used distributions in statistics. There are many normal related distribution such as Student's t distribution or chi-squared distributions.

Definition 43. Let $Z, Z_1, Z_2, \dots$ be a sequence of independent standard normal distributions, that is, $Z_i \sim i.i.d. N(0,1)$. The distribution of $U = Z^2$ is called chi-squared distribution with 1 degree of freedom, denoted by $U \sim \chi^2(1)$. Similarly $V = Z_1^2 + \cdots + Z_k^2$ is distributed from chi-squared with k degrees of freedom, denoted by $\chi^2(k)$.

Theorem 89. $\chi^2(1) \sim \text{gamma}(1/2, 1/2)$, $\chi^2(2) \sim \text{exponential}(1/2)$ and $\chi^2(k) \sim \text{gamma}(k/2, 1/2)$.

Proof. Suppose $Z_i \sim i.i.d. N(0,1)$. Then $V = Z_1^2 + \cdots + Z_k^2 \sim \chi^2(k)$ by definition. Consider $k = 1$ case first. The moment generating function of Z_1^2 is given by

$$\text{mgf}_{Z_1^2}(t) = \mathbb{E}[e^{tZ_1^2}] = \int e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \int e^{-z^2(1-2t)/2}\, dz = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi/(1-2t)} = (1-2t)^{-1/2}$$
$$\sim \text{gamma}(1/2, 1/2).$$

For general k, the moment generating function of V is

$$\text{mgf}_V(t) = \mathbb{E}[e^{tV}] = \mathbb{E}[e^{t(Z_1^2+\cdots+Z_k^2)}] = \mathbb{E}[e^{tZ_1^2}] \cdots \mathbb{E}[e^{tZ_k^2}] = \{\mathbb{E}[e^{tZ_1^2}]\}^k = (1-2t)^{-k/2} \sim \text{gamma}(k/2, 1/2).$$

If $k = 2$, $V \sim \text{gamma}(2/2, 1/2) \sim \text{gamma}(1, 1/2) \sim \text{exponential}(1/2)$. $\square$

Definition 44. Suppose $Z \sim N(0,1)$ and $V \sim \chi^2(k)$ are independent. The distribution of $T = Z/\sqrt{V/k}$ is called the Student's t distribution with k degrees of freedom.

Consider a transformation $g(z, v) = (z/\sqrt{v/k}, v)$. So that $z = t\sqrt{v/k}$ and the Jacobean is the determinant of

$$\begin{pmatrix} \sqrt{v/k} & t/2\sqrt{vk} \\ 0 & 1 \end{pmatrix},$$

that is, $\sqrt{v/k}$. The joint density of (t, v) is

$$\begin{aligned}
\mathrm{pdf}_{T,V}(t,v) &= \mathrm{pdf}_{Z,V}(t\sqrt{v/k}, v)\sqrt{v/k} = \frac{1}{\sqrt{2\pi}}\exp(-(t\sqrt{v/k})^2/2)(\Gamma(k/2)2^{k/2})^{-1}v^{k/2-1}\exp(-v/2)\sqrt{v/k} \\
&= \frac{1}{\sqrt{k\pi}\Gamma(k/2)2^{(k+1)/2}}v^{(k+1)/2-1}\exp(-v(t^2/k+1)/2).
\end{aligned}$$

The marginal density of T is given by

$$\begin{aligned}
\mathrm{pdf}_T(t) &= \int_0^\infty \mathrm{pdf}_{T,V}(t,v)\,dv = \frac{1}{\sqrt{k\pi}\Gamma(k/2)2^{(k+1)/2}}\int_0^\infty v^{(k+1)/2-1}\exp(-v(t^2/k+1)/2)\,dv \\
&= \frac{1}{\sqrt{k\pi}\Gamma(k/2)2^{(k+1)/2}}\frac{\Gamma((k+1)/2)}{((1+t^2/k)/2)^{(k+1)/2}} \\
&= \frac{\Gamma((k+1)/2)}{\sqrt{k\pi}\Gamma(k/2)}\Big(1+\frac{t^2}{k}\Big)^{-(k+1)/2}.
\end{aligned}$$

Example 103. The expectation and variance of t-distribution $T \sim \mathrm{t}(\nu)$ are

$$\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}\Big[\frac{Z}{\sqrt{U/\nu}}\Big] = \mathbb{E}[Z]\mathbb{E}[U^{-1/2}\nu^{1/2}] = 0 \\
\mathbb{V}\mathrm{ar}(T) &= \mathbb{E}[T^2] = \mathbb{E}[Z^2]\mathbb{E}[U^{-1}\nu] = 1\cdot\frac{\Gamma(\nu/2-1)\nu}{\Gamma(\nu/2)(1/2)^{-1}} = \frac{\nu}{\nu-2}
\end{aligned}$$

In general, kth moment exists for $k < \nu$. If k is odd, Z^k is odd and the density is even function, hence, $\mathbb{E}(T^k) = 0$. If k is even,

$$\mathbb{E}(T^k) = \mathbb{E}(Z^k)\mathbb{E}(U^{-k/2}\nu^{k/2}) = \frac{\Gamma(1/2+k/2)}{\Gamma(1/2)(1/2)^{k/2}}\frac{\Gamma(\nu/2-k/2)\nu^{k/2}}{\Gamma(\nu/2)(1/2)^{-k/2}} = \frac{(k-1)(k-3)\cdots 1\times\nu^{k/2}}{(\nu-2)(\nu-4)\cdots(\nu-k)}.$$

Thus $\mathbb{E}(T^3) = 0$ for $\nu > 3$, $\mathbb{E}(T^4) = \frac{3\nu^2}{(\nu-2)(\nu-4)}$ for $\nu > 4$ which implies the skewness is $\mathbb{E}[((T-\mu)/\sigma)^3] = \mathbb{E}[T^3]/\sigma^3 = 0$ and kurtosis $\mathbb{E}[((T-\mu)/\sigma)^4] = \mathbb{E}[T^4]/\sigma^4 = [3\nu^2/((\nu-2)(\nu-4))][\nu/(\nu-2)]^2 = 3(\nu-2)/(\nu-4) = 3+6/(\nu-4) > 3$. Hence Student's t-distribution has heavier tail than normal.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. normal with mean μ and variance σ^2, that is, $X_i \sim i.i.d.\ N(\mu,\sigma^2)$. Then

$$\begin{aligned}
\overline{X}_n &= \frac{1}{n}(X_1+\cdots+X_n) \xrightarrow{p} \mu \\
S_n^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X}_n)^2 = \frac{n}{n-1}\Big[\frac{1}{n}\sum_{i=1}^n X_i^2 - \big(\overline{X}_n\big)^2\Big] \xrightarrow{p} \mathbb{E}[X_1^2] - \big(\mathbb{E}[X_1]\big)^2 = \mathbb{V}\mathrm{ar}(X_1) = \sigma^2
\end{aligned}$$

Theorem 90. Assume $X_i \sim i.i.d.\ N(\mu,\sigma^2)$. $\overline{X}_n$ and $(X_1-\overline{X}_n, \ldots, X_n-\overline{X}_n)$ are independent.

Proof. Let $V = (V_1,\ldots,V_n) = (X_1-\overline{X}_n),\ldots,X_n-\overline{X}_n)$. The joint moment generating function is

$$\mathrm{mgf}_{\overline{X}_n,V}(s,t) = \mathrm{mgf}_{\overline{X}_n,V_1,\ldots,V_n}(s,t_1,\ldots,t_n) = \mathbb{E}[e^{s\overline{X}_n+\sum_{j=1}^n t_j(X_j-\overline{X}_n)}].$$

The exponent terms are

$$s\overline{X}_n + \sum_{j=1}^n t_j(X_j - \overline{X}_n) = \sum_{j=1}^n t_j X_j + (s - t_1 - \cdots - t_n)\overline{X}_n = \sum_{j=1}^n t_j X_j + (s - n\bar{t})\frac{1}{n}\sum_{j=1}^n X_j = \sum_{j=1}^n (s/n + t_j - \bar{t})X_j.$$

The independence of X_j's imply

$$\begin{aligned}
\mathrm{mgf}_{\overline{X}_n, V}(s,t) &= \mathbb{E}[e^{\sum_{j=1}^n (s/n+t_j-\bar{t})X_j}] = \prod_{j=1}^n \mathrm{mgf}_{X_j}(s/n + t_j - \bar{t}) = \prod_{j=1}^n e^{(s/n+t_j-\bar{t})\mu + (\sigma^2/2)(s/n+t_j-\bar{t})^2} \\
&= \exp\Big[\sum_{j=1}^n \Big((\frac{s}{n} + t_j - \bar{t})\mu + \frac{\sigma^2}{2}(\frac{s}{n} + t_j - \bar{t})^2\Big)\Big] \\
&= \exp\Big[s\mu + + \frac{\sigma^2}{2}\Big(n\frac{s^2}{n^2} + 2\frac{s}{n}\sum_{j=1}^n (t_j - \bar{t}) + \sum_{j=1}^n (t_j - \bar{t})^2\Big)\Big] \\
&= \exp\Big[s\mu + s^2\sigma^2/(2n)\Big] \times \exp\Big[\frac{\sigma^2}{2}\sum_{j=1}^n (t_j - \bar{t})^2\Big].
\end{aligned}$$

Hence $\overline{X}_n \sim N(\mu, \sigma^2/n)$ and $V \sim$ multivariate normal with mean 0 and variance $\sigma^2(I_n - \mathbf{1}_n\mathbf{1}_n^\top/n)$ are independent where I_n is diagonal $n \times n$ matrix with unit diagonal and $\mathbf{1}_n$ is $n \times 1$ column vector of 1's. $\square$

Note multivariate normal distributions $N(\mu, \Sigma)$ have densities

$$|2\pi\Sigma|^{-1/2} \exp\Big(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\Big)$$

and moment generating functions

$$\exp\Big(\mu^\top \mathbf{t} + \frac{1}{2}\mathbf{t}^\top \Sigma \mathbf{t}\Big).$$

Theorem 91. Assume $X_i \sim$ *i.i.d.* $N(\mu, \sigma^2)$. Then $\overline{X}_n$ and S_n^2 are independent and $\overline{X}_n \sim N(\mu, \sigma^2/n)$, $(n-1)S_n^2/\sigma^2 \sim \chi^2(n-1)$.

Proof. Since S_n^2 is a function of $V = (X_1 - \overline{X}_n, \ldots, X_n - \overline{X}_n)$, two random variables $\overline{X}_n$ and S_n^2 are independent. The fact $\overline{X}_n \sim N(\mu, \sigma^2/n)$ implies $\sqrt{n}(\overline{X}_n - \mu)/\sigma \sim N(0,1)$ and $n(\overline{X}_n - \mu)^2/\sigma^2 \sim \chi^2(1)$. Using $(X_j - \mu)/\sigma \sim N(0,1)$, $\frac{1}{\sigma^2}\sum_{j=1}^n (X_j - \mu)^2 \sim \chi^2(n)$ and the sum of squares becomes

$$\begin{aligned}
\sum_{j=1}^n (X_j - \mu)^2 &= \sum_{j=1}^n [(X_j - \overline{X}_n)^2 - 2(X_j - \overline{X}_n)(\overline{X}_n - \mu) + (\overline{X}_n - \mu)^2] = \sum_{j=1}^n (X_j - \overline{X}_n)^2 + n(\overline{X}_n - \mu)^2 \\
&= (n-1)S_n^2 + n(\overline{X}_n - \mu)^2.
\end{aligned}$$

Since $W = \frac{1}{\sigma^2}\sum_{j=1}^n (X_j - \overline{X}_n)^2$ and $n(\overline{X}_n - \mu)^2/\sigma^2 \sim \chi^2(1)$ are independent, the distribution of W can be determined using the moment generating function, that is,

$$\begin{aligned}
(1-2t)^{-n/2} &= \mathrm{mgf}_{\frac{1}{\sigma^2}\sum_{j=1}^n (X_j-\mu)^2}(t) = \mathbb{E}\Big[\exp\Big\{t\Big(\frac{1}{\sigma^2}\sum_{j=1}^n (X_j - \overline{X}_n)^2 + n(\overline{X}_n - \mu)^2/\sigma^2\Big)\Big\}\Big] \\
&= \mathrm{mgf}_W(t)\mathrm{mgf}_{n(\overline{X}_n-\mu)^2/\sigma^2}(t) = \mathrm{mgf}_W(t)(1-2t)^{-1/2}
\end{aligned}$$

implies $\text{mgf}_W(t) = (1-2t)^{-n/2}/(1-2t)^{-1/2} = (1-2t)^{-(n-1)/2} \sim \chi^2(n-1)$. $\square$

Theorem 92. Assume $X_i \sim i.i.d.\ N(\mu, \sigma^2)$. Then

$$\frac{\overline{X}_n - \mu}{\sqrt{S_n^2/n}} \sim t(n-1)$$

Proof. Note that $\sqrt{n}(\overline{X}_n - \mu)/\sigma \sim N(0,1)$ and $(n-1)S_n^2/\sigma^2 \sim \chi^2(n-1)$ are independent. Hence

$$\frac{\overline{X}_n - \mu}{\sqrt{S_n^2/n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S_n^2}{\sigma^2}/(n-1)} \sim \frac{N(0,1)}{\sqrt{\chi^2(n-1)/(n-1)}} \sim t(n-1).$$

The last distribution is the definition of t-distribution. $\square$