

556: MATHEMATICAL STATISTICS I

GENERAL RESULTS FOR THE SAMPLE MEAN AND VARIANCE STATISTICS

THEOREM

Suppose that X_1, \dots, X_n is a random sample from a distribution, say with finite expectation μ and variance σ^2 . Consider the sample mean and sample variance statistics \bar{X} and s^2 and denote

$$T_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad T_2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Then

- (a) $\mathbb{E}_{f_{T_1}}[T_1] = \mu$
- (b) $\text{Var}_{f_{T_1}}[T_1] = \frac{\sigma^2}{n}$
- (c) $\mathbb{E}_{f_{T_2}}[T_2] = \sigma^2$

Proof (a) and (b) follow from elementary properties of expectations and variances for independent random variables. For (c), note that

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

Hence

$$\begin{aligned} \mathbb{E}_{f_{T_2}}[T_2] &= \frac{1}{n-1} \mathbb{E}_{f_{\mathcal{X}}} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n \mathbb{E}_{f_{\mathcal{X}}}[X_i^2] - n \mathbb{E}_{f_{\mathcal{X}}}[\bar{X}^2] \right] \\ &= \frac{1}{n-1} \left[n(\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) \right] \\ &= \sigma^2 \end{aligned} \tag{1}$$

where line (1) follows from the fact that for any random variable X

$$\sigma^2 = \mathbb{E}_{f_X}[X^2] - \mathbb{E}_{f_X}[X]^2 = \mathbb{E}_{f_X}[X^2] - \mu^2$$

and the result of parts (a) and (b).

SAMPLING FROM A NORMAL FAMILY

Recall the fundamental transformation results for Normal random variables:

(i) If $X \sim \mathcal{N}(0, 1)$, then

$$X^2 \sim \chi_1^2 \equiv \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

(ii) If $X_1, \dots, X_r \sim \mathcal{N}(0, 1)$ are independent random variables, then

$$Y = \sum_{i=1}^r X_i^2 \sim \chi_r^2 \equiv \text{Gamma}\left(\frac{r}{2}, \frac{1}{2}\right)$$

(iii) If $Y_1 \sim \chi_{r_1}^2$ and $Y_2 \sim \chi_{r_2}^2$ are independent random variables, then

$$Y = Y_1 + Y_2 \sim \chi_{r_1+r_2}^2$$

THEOREM

Suppose that X_1, \dots, X_n is a random sample from a normal distribution, say $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Define the sample mean and sample variance statistics \bar{X} and s^2 as the random variables

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

Then

(a) $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$

(b) \bar{X} is independent of $\{X_i - \bar{X}, i = 1, \dots, n\}$, and \bar{X} and s^2 are independent random variables

(c) The random variable

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

has a **chi-squared distribution** with $n-1$ degrees of freedom.

Proof (a) Proof straightforward using mgfs.

(b) The joint pdf X_1, \dots, X_n is the normal density

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}$$

Consider the multivariate transformation to Y_1, \dots, Y_n where

$$\left. \begin{array}{l} Y_1 = \bar{X} \\ Y_i = X_i - \bar{X}, i = 2, \dots, n \end{array} \right\} \iff \left\{ \begin{array}{l} X_1 = Y_1 - \sum_{i=2}^n Y_i \\ X_i = Y_i + Y_1, i = 2, \dots, n \end{array} \right.$$

Thus $\underline{Y} = A\underline{X}$, or equivalently $\underline{X} = A^{-1}\underline{Y}$, where A is the $n \times n$ matrix with (i, j) th element

$$[A]_{ij} = \begin{cases} \frac{1}{n} & i = 1, j = 1, 2, \dots, n \\ 1 - \frac{1}{n} & i = j = 2, 3, \dots, n \\ -\frac{1}{n} & \text{otherwise} \end{cases}$$

that is, we have a linear transformation. Note that

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^n \left[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - \mu) + (\bar{x} - \mu)^2 \right] \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \end{aligned}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ is the observed sample mean. Thus the joint pdf of X_1, \dots, X_n takes the form

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right\}.$$

Now

$$x_1 - \bar{x} = -\sum_{i=2}^n (x_i - \bar{x}) = -\sum_{i=2}^n y_i$$

and so

$$\sum_{i=1}^n (x_i - \bar{x})^2 = (x_1 - \bar{x})^2 + \sum_{i=2}^n (x_i - \bar{x})^2 = \left(-\sum_{i=2}^n y_i \right)^2 + \sum_{i=2}^n y_i^2$$

The Jacobian of the transformation is n , so the joint density of Y_1, \dots, Y_n is given by the multivariate transformation theorem as

$$\begin{aligned} f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) &= n \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[\left(-\sum_{i=2}^n y_i \right)^2 + \sum_{i=2}^n y_i^2 + n(y_1 - \mu)^2 \right] \right\} \\ &= n \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[\left(-\sum_{i=2}^n y_i \right)^2 + \sum_{i=2}^n y_i^2 \right] \right\} \times \exp \left\{ -\frac{n}{2\sigma^2} (y_1 - \mu)^2 \right\} \\ &= f_{Y_2, \dots, Y_n}(y_2, \dots, y_n) f_{Y_1}(y_1) \end{aligned}$$

and therefore Y_1 is independent of Y_2, \dots, Y_n . Hence \bar{X} is **independent** of the random variables $\{Y_i = X_i - \bar{X}, i = 2, \dots, n\}$. Finally, \bar{X} is also independent of $X_1 - \bar{X}$ as

$$X_1 - \bar{X} = -\sum_{i=2}^n (X_i - \bar{X})$$

and of s^2 , which is a function only of $\{X_i - \bar{X}, i = 1, \dots, n\}$. As \bar{X} is independent of these variables, \bar{X} and s^2 are also independent.

(c) The random variables that appear as sums of squares terms that joint pdf are

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

or $V_1 = V_2 + V_3$, say. Now, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, so therefore

$$\frac{(X_i - \mu)^2}{\sigma^2} \sim \mathcal{N}(0, 1) \implies \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_1^2 \equiv \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) \implies V_1 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

as the X_i s are independent, and the sum of n independent $Ga(1/2, 1/2)$ variables has a $Ga(n/2, 1/2)$ distribution. Similarly, as $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, $V_3 \sim \chi_1^2$. By part (b), V_2 and V_3 are independent, and so the mgfs of V_1 , V_2 and V_3 are related by

$$M_{V_1}(t) = M_{V_2}(t)M_{V_3}(t) \implies M_{V_2}(t) = \frac{M_{V_1}(t)}{M_{V_3}(t)}$$

As V_1 and V_3 are Gamma random variables, M_{V_1} and M_{V_3} are given by

$$M_{V_1}(t) = \left(\frac{1/2}{1/2 - t}\right)^{n/2} \quad \text{and} \quad M_{V_3}(t) = \left(\frac{1/2}{1/2 - t}\right)^{1/2}.$$

So therefore

$$M_{V_2}(t) = \left(\frac{1/2}{1/2 - t}\right)^{(n-1)/2}$$

which is also the mgf of a Gamma random variable, and hence

$$V_2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

and the result follows.

Alternative inductive proof of (c): Let \bar{X}_k and s_k^2 , $k = 1, 2, \dots, n$ denote the sample mean and sample variance random variables derived from the first k variables. Now, for $k \geq 2$, it can be shown after some manipulation that

$$(k-1)s_k^2 = (k-2)s_{k-1}^2 + \left(\frac{k-1}{k}\right)(X_k - \bar{X}_{k-1})^2 \quad (2)$$

For $k = 2$

$$(2-1)s_2^2 = \frac{1}{2}(X_2 - X_1)^2 = \left(\frac{X_2 - X_1}{\sqrt{2}}\right)^2 = Z^2$$

say, where $Z \sim \mathcal{N}(0, 1)$. Thus $s_2^2 \sim \chi_1^2$. Now for the inductive hypothesis, presume that

$$(k-1)s_k^2 \sim \chi_{k-1}^2$$

so that, using the identity in (2),

$$ks_{k+1}^2 = (k-1)s_k^2 + \left(\frac{k}{k+1}\right)(X_{k+1} - \bar{X}_k)^2$$

The two terms on the right hand side are independent (using the result in (b)); the first term is χ_{k-1}^2 distributed, the second term is χ_1^2 distributed, so ks_{k+1}^2 is χ_k^2 distributed and the inductive argument is completed.