STK4011 and STK9011

Autumn 2016

Properties of a random sample

Covers (most of) the following material from chapter 5:

- Section 5.1: pages 207-208
- Section 5.2: pages 211-216
- Section 5.3: pages 218-220

Ørnulf Borgan
Department of Mathematics
University of Oslo

Random sample

The random variables $X_1, X_2, \ldots, X_n$ are called a random sample from the population $f(x)$ if $X_1, X_2, \ldots, X_n$ are mutually independent and the marginal pmf or pdf of each X_i is $f(x)$

Alternatively, we may say that $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) random variables with pmf or pdf $f(x)$

The joint pmf or pdf is

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Statistics and sampling distributions

A function $Y = T(X_1, X_2, \ldots, X_n)$ of $X_1, X_2, \ldots, X_n$ is called a statistic

A statistic may be real-valued or vector-valued

The probability distribution of a statistic Y is called the sampling distribution of Y

Two common examples of statistics:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{(sample mean)}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \qquad \text{(sample variance)}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2

It is well known that (cf. page 214)

$$\mathrm{E}\bar{X} = \mu$$

$$\mathrm{Var}\,\bar{X} = \frac{\sigma^2}{n}$$

$$\mathrm{E}S^2 = \sigma^2$$

Thus $\bar{X}$ is an unbiased estimator of μ and S^2 is an unbiased estimator of σ^2

The distribution of the sample mean

We will have a closer look at the sampling distribution of $\bar{X}$

The mgf of $\bar{X}$ is given by $M_{\bar{X}}(t) = [M_X(t/n)]^n$

where $M_X(t)$ is the mgf of $X_1, X_2, ..., X_n$

If we have a random sample from a $n(\mu, \sigma^2)$ population, we have $M_X(t) = \exp\{\mu t + \sigma^2 t^2 / 2\}$

Then

$$M_{\bar{X}}(t) = [M_X(t/n)]^n = \left[\exp\{\mu(t/n) + \sigma^2 (t/n)^2 / 2\}\right]^n$$

$$= \exp\left\{n\left[\mu(t/n) + \sigma^2 (t/n)^2 / 2\right]\right\} = \exp\{\mu t + (\sigma^2 / n) t^2 / 2\}$$

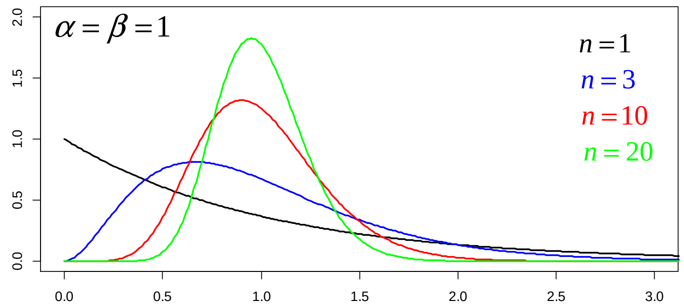
It follows that $\bar{X} \sim n(\mu, \sigma^2 / n)$

If we have a random sample from a $\text{gamma}(\alpha, \beta)$ population, we have $M_X(t) = (1/(1-\beta t))^\alpha$

Then

$$M_{\bar{X}}(t) = [M_X(t/n)]^n = \left[(1/(1-\beta(t/n)))^\alpha\right]^n = (1/(1-(\beta/n)t))^{n\alpha}$$

It follows that $\bar{X} \sim \text{gamma}(n\alpha, \beta/n)$



When one cannot use mgfs to find the distribution of sums and averages, one has to resort to the convolution formula:

If X and Y are independent continuous random variables with pdfs $f_X(x)$ and $f_Y(x)$, the pdf of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\, dw$$

To prove the result, one first finds the joint pdf of $Z = X + Y$ and W. Integrating out w, one then finds the marginal pdf of Z

Example: Sum of Cauchy random variables

Let U and V be independent Cauchy random variables, $U \sim \text{Cauchy}(0, \sigma)$ and $V \sim \text{Cauchy}(0, \tau)$, i.e.

$$f_U(u) = \frac{1}{\pi\sigma} \frac{1}{1 + (u/\sigma)^2} \qquad -\infty < u < \infty$$

$$f_V(u) = \frac{1}{\pi\tau} \frac{1}{1 + (v/\tau)^2} \qquad -\infty < v < \infty$$

By the convolution formula we have that

$$f_Z(z) = \int_{-\infty}^{\infty} f_U(w) f_V(z-w)\, dw$$

$$= \int_{-\infty}^{\infty} \frac{1}{\pi\sigma} \frac{1}{1 + (w/\sigma)^2} \frac{1}{\pi\tau} \frac{1}{1 + ((z-w)/\tau)^2}\, dw$$

The integral is quite «tricky», but by using integration by partial fractions one may show that (cf. exercise 5.7)

$$f_Z(z) = \frac{1}{\pi(\sigma+\tau)} \frac{1}{1+(z/(\sigma+\tau))^2}$$

Thus $Z \sim \text{Cauchy}(0, \sigma+\tau)$

From this it follows that if $Z_1, ..., Z_n$ are iid $\text{Cauchy}(0,1)$, then $\sum Z_i \sim \text{Cauchy}(0, n)$ and $\bar{Z} \sim \text{Cauchy}(0,1)$

The sample mean has the same distribution as the individual observations!

Sampling from the normal distribution

Let $X_1, X_2, ..., X_n$ be a random sample from a $n(\mu, \sigma^2)$ distribution

Then

a) $\bar{X}$ and S^2 are independent

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

c) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

We will show a)

It is sufficient to prove the result for $\mu = 0$ and $\sigma = 1$

Proof of a):

Note that we may write

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

$$= \frac{1}{n-1}\left\{\left(X_1 - \bar{X}\right)^2 + \sum_{i=2}^{n}\left(X_i - \bar{X}\right)^2\right\}$$

$$= \frac{1}{n-1}\left\{\left[\sum_{i=2}^{n}\left(X_i - \bar{X}\right)\right]^2 + \sum_{i=2}^{n}\left(X_i - \bar{X}\right)^2\right\}$$

Thus S^2 is a function of $(X_2 - \bar{X}, \text{ .... }, X_n - \bar{X})$

It is therefore sufficient to show that $\bar{X}$ and the random vector $(X_2 - \bar{X}, \text{ .... }, X_n - \bar{X})$ are independent

The joint pdf of $X_1, X_2, ..., X_n$ is given by (when $\mu = 0$ and $\sigma = 1$)

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right)$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$$

We now make the transformation:

$$Y_1 = \bar{X}, \; Y_2 = X_2 - \bar{X}, \text{ .... }, Y_n = X_n - \bar{X}$$

The inverse transformation is given by

$$X_1 = Y_1 - \sum_{i=2}^{n} Y_i$$
$$X_n = Y_i + Y_1 \quad \text{for } i = 2,....,n$$

The Jacobian becomes

$$J(y_1,...,y_n) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} 1 & -1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

We may show that the Jacbobian equals n

Hence the joint pdf of $Y_1,....,Y_n$ is given by

$$f(y_1,....,y_n) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left(y_1 - \sum_{i=2}^{n} y_i\right)^2 - \frac{1}{2}\sum_{i=2}^{n}(y_i + y_1)^2\right\} \cdot n$$

$$= \left(\frac{n}{2\pi}\right)^{1/2} \exp\left(-\frac{n}{2}y_1^2\right) \frac{n^{1/2}}{(2\pi)^{(n-1)/2}} \exp\left\{-\frac{1}{2}\sum_{i=2}^{n} y_i^2 - \frac{1}{2}\left(\sum_{i=2}^{n} y_i\right)^2\right\}$$

Since the joint pdf factors, it follows that Y_1 and $(Y_2,....,Y_n)$ are independent

Now $\bar{X}$ is a function of Y_1 and S^2 is a function of $(Y_2,....,Y_n)$, so $\bar{X}$ and S^2 are independent

t-distribution

Let $X_1, X_2,...., X_n$ be a random sample from the $n(\mu,\sigma^2)$ distribution

Then $U = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim n(0,1)$ and $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

and the two variables are independent

Note that

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ (n-1)}} = \frac{U}{\sqrt{V/(n-1)}}$$

T is t-distributed with $df = n - 1$

If $U \sim n(0,1)$ and $V \sim \chi^2_p$ are independent, then

$$T = \frac{U}{\sqrt{V/p}}$$

is t-distributed with $df = p$. The pdf is given by

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{\left(1 + \frac{t^2}{p}\right)^{(p+1)/2}} \qquad -\infty < t < \infty$$

To prove the result, one first finds the joint pdf of $T = U / \sqrt{V / p}$ and $W = V$. Integrating out w, one then finds the marginal pdf of T

For $p = 1$ we have the Cauchy distribution