The sample mean and variance

September 12, 2013

The sample mean and variance

Sample mean and sample variance

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with expected value μ and variance σ^2.

Consider the sample mean and the sample variance:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

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

Sample mean and sample variance

We have

$$\mathrm{E}\left(\overline{X}\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) = \mu$$

$$\begin{aligned}
\mathrm{Var}\left(\overline{X}\right) &= \mathrm{E}\left(\left(\overline{X} - \mu\right)^2\right) \\
&= \frac{1}{n^2}\,\mathrm{E}\left(\sum_{i=1}^{n}\sum_{i=j}^{n} (X_i - \mu)\,(X_j - \mu)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) + \frac{1}{n^2}\sum\sum_{i \neq j} \mathrm{Cov}(X_i, X_j) = \frac{\sigma^2}{n}
\end{aligned}$$

Sample mean and sample variance

Moreover,

$$(n-1)\,S^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i - \mu + \mu - \overline{X})^2$$

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

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

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

Sample mean and sample variance

Thus,

$$\mathrm{E}\left((n-1)\,S^2\right) = \sum_{i=1}^{n} \mathrm{E}\left((X_i - \mu)^2\right) - n\,\mathrm{E}\left((\overline{X} - \mu)^2\right)$$

$$= n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$

Therefore

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

- S^2 is an unbiased estimator of σ^2.

Sample mean and sample variance

Moreover,

$$\mathrm{Cov}\left(\overline{X}, X_i - \overline{X}\right) = \mathrm{Cov}\left(\overline{X}, X_i\right) - \mathrm{Var}\left(\overline{X}\right)$$

$$= \frac{1}{n}\,\mathrm{Cov}\left(X_i + \sum_{j \neq i} X_j, X_i\right) - \frac{\sigma^2}{n}$$

$$= \frac{1}{n}\,\mathrm{Var}(X_i) - \frac{\sigma^2}{n} = 0$$

since $\mathrm{Cov}(X_i, X_j) = 0$ for $i \neq j$.

- $\overline{X}$ and each $X_i - \overline{X}$ are uncorrelated r.v.

The gaussian case

Suppose now that the random variables $X_1, X_2, \ldots, X_n$ are gaussian.

- The sample mean $\overline{X}$ is gaussian,

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

- Since $X_i - \overline{X}$ and $\overline{X}$ are uncorrelated r.v. and $X_i - \overline{X}$ is also gaussian, $X_i - \overline{X}$ and $\overline{X}$ are independent r.v.
- Hence, $\overline{X}$ and S^2 are (in the gaussian case) independent r.v.
- Which is the probability law of S^2?

The χ^2-distribution

Definition

Let $Z_1, Z_2, \ldots, Z_n$ be standard normal $\mathrm{N}(0,1)$ independent r.v. The r.v.

$$\chi^2(n) = Z_1^2 + \cdots + Z_n^2$$

is said a chi-squared r.v. with n degrees of freedom.

The χ^2-distribution

Let us calculate the moment generating function of each Z_i^2.

$$\mathrm{E}\left(e^{t\,Z_i^2}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t\,x^2}\, e^{-x^2/2}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2(1-2t)/2}\, dx = \frac{1}{\sqrt{1-2t}}, \quad t < \frac{1}{2}$$

Hint: This result can be obtained letting $\sigma^2 = 1/(1-2t) > 0$ and noticing that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/(2\sigma^2)}\, dx = \sigma$$

The χ^2-distribution

Therefore the moment generating function of the $\chi^2(n)$ r.v. is

$$\phi(t) = \mathrm{E}\left(e^{t\,\sum_{i=1}^n Z_i^2}\right)$$

$$= \prod_{i=1}^{n} \mathrm{E}\left(e^{t\,Z_i^2}\right) = \frac{1}{(1-2t)^{n/2}}, \quad t < \frac{1}{2}$$

The distribution of the sample variance

Since

$$(n-1)\,S^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X} - \mu)^2$$

we have

$$\frac{(n-1)\,S^2}{\sigma^2} + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

The distribution of the sample variance

Notice that

- $$\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$
  (one squared standard normal)

- $$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$$
  (sum of n independent squared standard normal)

The distribution of the sample variance

Taking generating functions and applying the convolution theorem,

$$\mathrm{E}\left(e^{t\,\frac{(n-1)\,S^2}{\sigma^2}}\right)\frac{1}{(1-2t)^{1/2}} = \frac{1}{(1-2t)^{n/2}}$$

Therefore

$$\mathrm{E}\left(e^{t\,\frac{(n-1)\,S^2}{\sigma^2}}\right) = \frac{1}{(1-2t)^{\frac{n-1}{2}}}$$

Hence,

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

The distribution of the sample variance

Theorem

Let $X_1, X_2, \ldots X_n$ be independent $\mathrm{N}(\mu,\sigma^2)$ random variables. Then $\overline{X}$ and S^2 are independent and

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

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

The χ^2-density

It can be proved that if $Z \sim \chi^2(n)$, then

$$f_Z(z) = \begin{cases} 0, & z < 0 \\ \frac{1}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}\right)^{\frac{n}{2}} z^{\frac{n}{2}-1}\, e^{-\frac{z}{2}}, & z > 0 \end{cases}, \quad n = 1, 2, 3, \ldots$$

- In particular, for $n = 2$ we have $Z \sim \mathrm{Exp}\left(\frac{1}{2}\right)$.

The χ^2-density

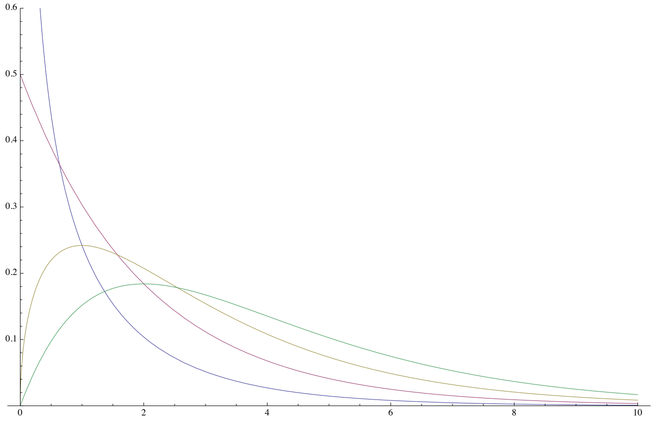

Figure: $f_Z(z)$ for $n = 1, 2, 3, 4$