Math 408 - Mathematical Statistics

Lecture 13-14. The Sample Mean and the Sample Variance Under Assumption of Normality

February 20, 2013

Framework

Let $X_1, \dots, X_n$ be a sample drawn from a population.

Suppose that the population is "Gaussian" $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

We want to estimate population parameters μ and σ^2.

Definition

- The **sample mean** is $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$
- The **sample variance** is $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$

Theorem

$\overline{X}_n$ and S_n^2 are **unbiased estimators** *of μ and σ^2, respectively,*

$$\mathbb{E}[\overline{X}_n] = \mu, \qquad \mathbb{E}[S_n^2] = \sigma^2$$

Our goal: to describe distributions of $\overline{X}_n$ and S_n^2

Distribution of $\overline{X}_n$

Theorem

If $X_1, \ldots, X_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables, then

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

Distribution of S_n^2

Theorem

If $X_1, \ldots, X_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables, then

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

The χ^2-distribution

Definition

Let $Z_1, \ldots, Z_n$ be independent standard normal variables,

$$Z_1, \ldots, Z_n \sim N(0,1)$$

Then the distribution of

$$Q = Z_1^2 + Z_2^2 + \ldots + Z_n^2$$

is called the χ^2**-distribution** with n **degrees of freedom**,

$$Q \sim \chi_n^2$$

- Probability Density Function:

$$\pi(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

  - $x \geq 0$
  - Γ is the gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$

The χ^2-distribution

The χ^2-distribution is especially important in hypothesis testing.

Nice Properties:

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

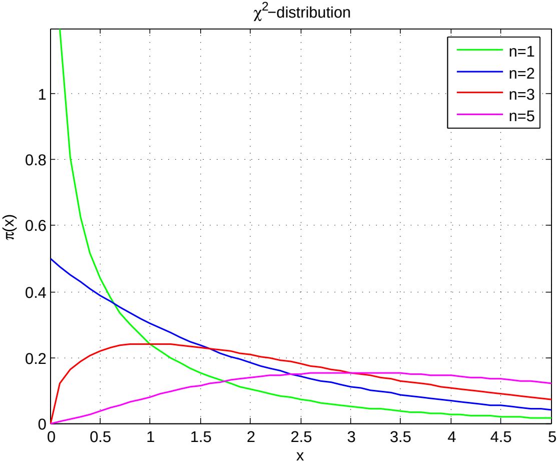
and

$$\left(\frac{X - \mu}{\sigma}\right)^2 \sim \chi_1^2$$

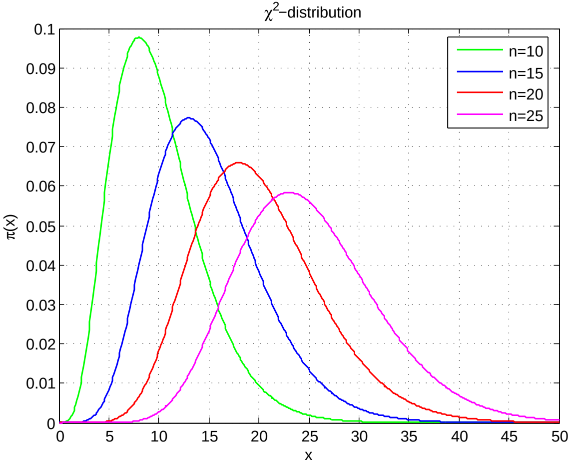
- If $U \sim \chi_n^2$ and $V \sim \chi_m^2$, and U and V are independent, then

$$U + V \sim \chi_{n+m}^2$$

Graph of the χ_n^2 PDF: small n



Graph of the χ_n^2 PDF: large n



- CLT: χ_n^2 converges to a normal distribution as $n \to \infty$
- $\chi_n^2 \to \mathcal{N}(n, 2n)$, as $n \to \infty$
- When $n > 50$, for many practical purposes, $\chi_n^2 = \mathcal{N}(n, 2n)$

Distribution of S_n^2

Theorem

If $X_1, \dots, X_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables, then

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Proof: is based on moment-generating functions...

Moment-generating functions

Definition

The moment-generating function (MGF) of a random variable $X \sim f(x)$ is

$$M(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

(if the expectation is defined)

Important Properties of MGFs:

- If $\exists\, \varepsilon > 0$ such that $M(t)$ exists for all $t \in (-\varepsilon, \varepsilon)$, then $M(t)$ uniquely determines the probability distribution, $M(t) \rightsquigarrow f(x)$.
- If $M(t)$ exists in an open interval containing zero, then

$$M^{(r)}(0) = \mathbb{E}[X^r] \qquad \text{(hence the name)}$$

To find moments $\mathbb{E}[X^r]$, we must do integration.
Knowing the MGF allows to replace integration by differentiation.

Moment-generating functions

Important Properties of MGFs: (continuation)

- If X has the MGF $M_X(t)$ and $Y = a + bX$, then

$$M_Y(t) = e^{at} M_X(bt)$$

- If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

- If X and Y have a joint distribution, then their joint MGF is defined as

$$M_{X,Y}(s,t) = \mathbb{E}[e^{sX+tY}]$$

  X and Y are independent if and only if

$$M_{X,Y}(s,t) = M_X(s) M_Y(t)$$

Moment-generating functions: Limitations and Examples

The major limitation of the moment-generating function is that it may not exist.

In this case, the characteristic function may be used:

$$\phi(t) = \mathbb{E}[e^{itX}]$$

Examples:

- $\mathcal{N}(\mu, \sigma^2)$:

$$M(t) = e^{\mu t} e^{\sigma^2 t^2/2}$$

- χ^2_n:

$$M(t) = (1 - 2t)^{-n/2}$$

Distribution of S_n^2

Theorem

If $X_1, \ldots, X_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables, then

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Bringing the t-distribution into the Game

Theorem

If $X_1, \ldots, X_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables, then

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

The t-distribution

Definition

Let $Z \sim \mathcal{N}(0,1)$, $U \sim \chi^2_n$, and Z and U are independent. Then the distribution of

$$T = \frac{Z}{\sqrt{U/n}}$$

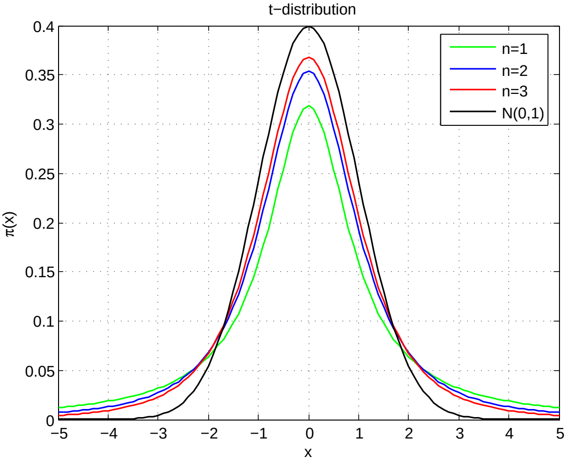
is called the t-**distribution** with n **degrees of freedom**.

- Probability Density Function:

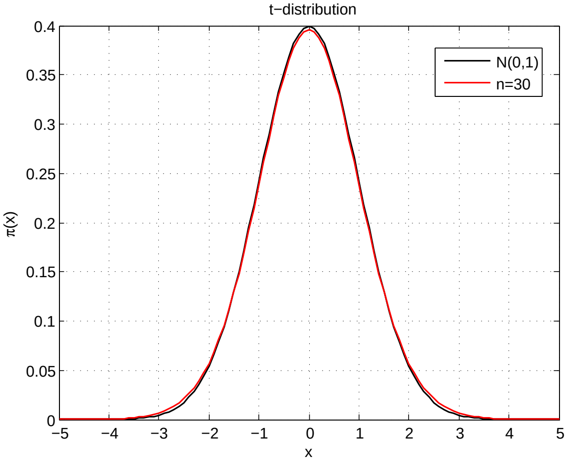
$$\pi(x) = \frac{\Gamma\left((n+1)/2\right)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

- The t-distribution is symmetric about zero, $\pi(x) = \pi(-x)$
- As $n \to \infty$, the t-distribution tends to the standard normal distribution. In fact, when $n > 30$, the two distributions are very close.

Graph of the t-distribution PDF: small n



Graph of the t-distribution PDF: large n



Bringing the t-distribution into the Game

Theorem

If $X_1, \ldots, X_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables, then

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

Summary

Under Assumption of Normality, $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$,

$$\text{the sample mean:} \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\text{the sample variance:} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

have the following properties:

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

- $$\boxed{\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}} \qquad \chi^2_n = \mathcal{N}(0,1)^2 + \ldots + \mathcal{N}(0,1)^2$$

- $$\boxed{\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}} \qquad t_n = \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2_n/n}}$$