Three common distributions

Miaoyan Wang

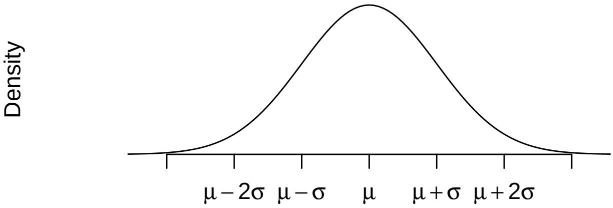
Department of Statistics
UW Madison

Normal Distribution

- **Normal distribution** is also known as a **Gaussian distribution**.
- The 1-d density function is given by the equation:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}$$

- μ and σ^2 are the **parameters** of the curve.
- $Y \sim N(\mu, \sigma^2)$ denotes that a random variable Y follows a normal distribution with mean μ and variance σ^2 (i.e. standard deviation σ).



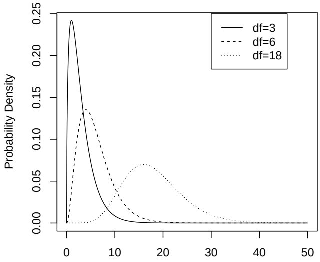
Chi-Squared Distribution

Definition of χ^2 distribution

Consider a random variable $V^2 = \sum_{j=1}^{m} Z_j^2$ with $Z_1, \ldots, Z_m$ being independent $N(0,1)$. Then, the distribution of V^2 is called a chi-square distribution with m degrees of freedom, denote by $V^2 \sim \chi^2_m$.

Properties of $V^2 \sim \chi^2_m$:

- The range of possible values of V^2 is from 0 to $+\infty$.
- The distribution is right-skewed.
- $E(V^2) = m$ and $Var(V^2) = 2m$.

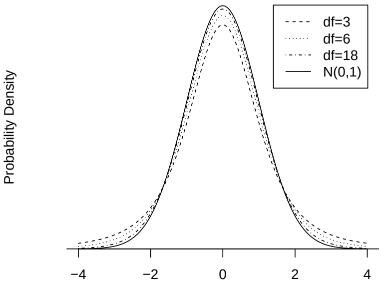


T distribution

Definition of T-distribution

Consider a random variable $V^2 = \sum_{j=1}^{m} Z_j^2$ with $Z_1, \ldots, Z_m$ being independent $N(0,1)$. Let X be another N(0,1) independent of $(Z_i)_{1 \leq j \leq m}$. Then, the ratio follows a t-distribution with m degrees of freedom,

$$\frac{X}{V^2} \sim T_{m-1}.$$



T Distribution

- The possible range of values of T is from $-\infty$ to ∞.
- The distribution is symmetric and centered around 0.
- The distribution of T is similar to that of Z (standard normal); probabilities are interpreted as areas under a probability density curve.
- However, the T-distribution has less concentration of probability close to 0 and more probability in the tails than Z.
- Also, the distribution of T depends on the sample size n.
- For larger n, the distribution curve has thinner tails, less spread, and more Z-like.

Sample Mean: central limit theorem

- Consider a random sample $Y_1, Y_2, \ldots, Y_n$, where Y_i's are independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2.
- Denote $Y_i \sim_{\text{i.i.d}} D(\mu, \sigma^2)$ for $i = 1, \ldots, N$. Note that the distribution D does not have to be Normal!

Central limit theorem (CLT)

Let $Y_1, \ldots, Y_n$ be i.i.d. sample from a common distribution with mean μ and sd σ. Define the sample mean $\bar{Y}(n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Then

$$\bar{Y}_n \overset{\text{in dist'n}}{\rightarrow} \mathcal{N}(\mu, \frac{\sigma^2}{n}), \quad \text{as} \quad n \rightarrow \infty.$$

The sample mean for a large i.i.d. sample is approximately $N(\mu, \frac{\sigma^2}{n})$. Why $\frac{\sigma^2}{n}$?

Sample Variance

- Again, consider $Y \sim N(\mu, \sigma^2)$.
- Let $Y_1, Y_2, \ldots, Y_n$ denote an i.i.d. sample from this population $N(\mu, \sigma^2)$.
- The sample variance is defined as

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

- As the sample size n increases, the sample variance gets closer to the population variance.
- Note that S^2 is also a random variable.

Properties of Sample Variance

- The expected value of the sample variance is

$$E\left(S^2\right) = \sigma^2$$

- Interpretation: Unbiased estimator
- The variance of the sample variance is

$$Var(S^2) = \frac{2\sigma^4}{n-1}.$$

- Interpretation: As n increases, variance decreases $\rightarrow$ precision to estimate σ^2 increases.

Standardization of Sample Variance

- Next goal: What is the distribution of S^2?
- Suppose $Y_1, Y_2, \ldots, Y_n$ is an i.i.d. sample from a normal distribution with mean μ and σ^2.
- Let S^2 be the sample variance.
- Define

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

- V^2 follows a chi-squared distribution with $n-1$ degrees of freedom.
- $\bar{Y}$ and S^2 are independent (next slide).

Independence between sample mean and sample variance

Basu's theorem

Suppose $Y_1, Y_2, \ldots, Y_n$ is an i.i.d. sample from $N(\mu, \sigma^2)$. Let $\bar{Y} = \frac{1}{n}\sum_i Y_i$ and $S^2 = \frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2$. Then $\bar{Y}$ and S^2 are independent.

Proof sketch (details as exercise)

Transform the random variables $\{Y_i\}$ to $\{X_i\}$, where

$$X_1 = \bar{Y}, \quad X_2 = Y_2 - \bar{Y}, \quad \ldots, \quad X_n = Y_n - \bar{Y}.$$

The joint density of $(X_1, \ldots, X_n)$ equals joint density of $(Y_1, \ldots, Y_n)$ times the Jacobian transformation (a constant). Straightforward calculations shows that

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = C \exp\left\{ -\frac{1}{2\sigma^2}\left(\sum_{i \geq 2} x_i\right)^2 + \sum_{i \geq 2} x_i^2 \right\} \exp\left\{ -\frac{n}{2\sigma^2}(x_1 - \mu)^2 \right\}.$$

The density factorizes w.r.t. x_1 and $(x_2, \ldots, x_n) \Longrightarrow \bar{Y}$ and S^2 are independent.