Summary of Dr. Kang's note on Probability Distributions

Discrete random variables

- Can only possess integer values (whole numbers). No decimal values.
- Possible values in the sample space can be counted or visualized.

Dr. Kang only discussed Bernoulli random variables (X = 0, 1) and Binomial random variables (X = 0, 1, 2, …,n) where n is the sample size but there are other types of discrete random variables such as: Negative Binomial, Poisson, Geometric, Hypergeometric random variables.
For any discrete random variable X, we have the following

- Probability mass function (pmf)

$$f(x) = P(X = x),\ x = \text{realization of X}$$

- Cumulative distribution function (Cdf)

$$F(x) = P(X \leq x)$$

Stata functions for Binomial distribution

- binomialp(n, x, p) for calculating pmf $f\ (x)\ =\ P(\ X =\ x)$

code: disp binomialp(n, x, p), n = sample size, x = realization of X and p = prob. of success

- binomial(n, x, p)for calculating $F\ (x)\ =\ P(\ X \leq\ x)$

code: disp binomial(n, x, p)

- comb(n, x) for calculating nCx

code: disp com(n,x)

- factorial(n) or n!

code: disp exp(lnfactorial(n))

Example 4.3.3 (Page 103). **We will practice this together with Stata.**
10% of a certain population is color blind. That is, p = 0.1

Draw a random sample of 25 people from the population and calculate the following probabilities

(a) Five or fewer will be color blind. That is: $P(X \leq x = 5) = disp\ binomial(25{,}5,.1)$.

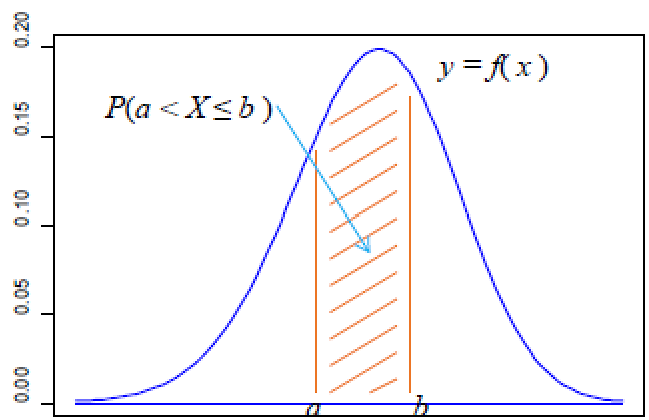
(b) Six or more will be color blind. That is: $P(X \geq 6 = x) = 1 - P(X \leq 5 = x - 1) = 1 = disp\ binomial(25{,}5,.1)$.

(c) Between six and nine inclusive will be color blind. That is: $P(6 \leq X \leq 9) = P(X \leq 9) - P(X \geq 6) = P(X \leq 9) - P(X \leq 5) = disp\ binomal(25{,}9,.1) - disp\ binomial(25{,}5,.1)$.

(d) Two, three or four will be color blind. That is: $P(X \leq 4) = disp\ binomial(25{,}4,.1)$

(e) At least one is color blind $P(X \geq 1) = 1 - P(X \leq 0) = 1 - P(X = 0) = disp\ 1 - binomialp(25{,}0,.1)$

Continuous random variables



- Can take any value within defined interval or range.

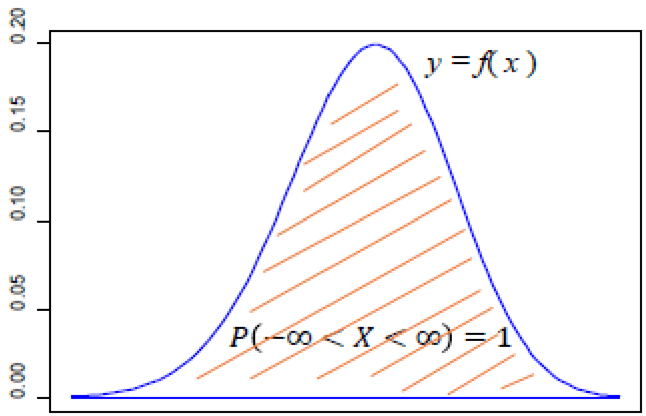
Examples of continuous random variable

X= the height of a randomly selected male physician from UNM Health Center.

R=rate of discovery of patients in a hospital.

S = salary of physicians in New Mexico

Probability density function (pdf)



$f(x)$ is called the probability density function (pdf) of X

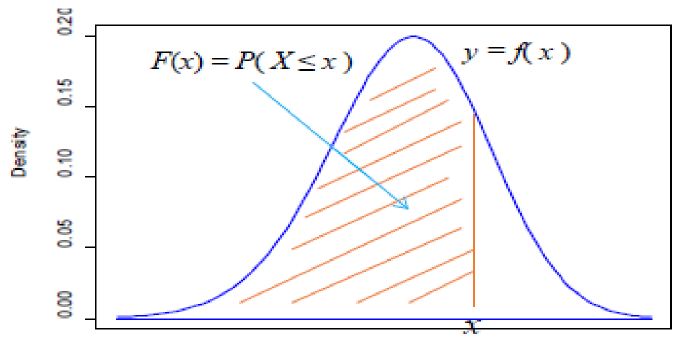
Properties

- $f(x) > 0$
- $P(X = c) = 0$ for any specific value c
- $P(a \leq X \leq b) = P(a < X < b)$
- $P\ (-\infty < X < \infty) = \ 1$, which is the entire area under the graph of $f(x)$ and above x-axis.

Cumulative distribution function (cdf)

- $F(x) = P(X \leq x)$ is the area enclosed by the graph of $f(x)$, x-axis and the vertical line at the point x. That is,

  $F(x) = P(X \leq x) = P(-\infty < X \leq x)$.



Facts

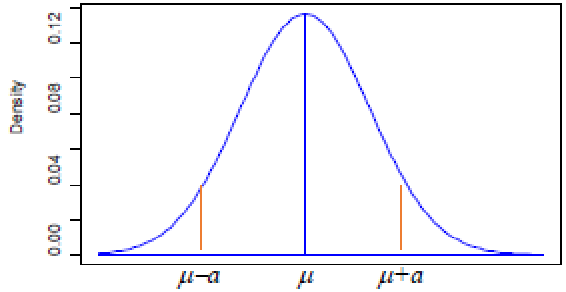
- $x_1 \leq x_2,\ \ then\ \ F(x_1) \leq F(x_2)$
- $\lim_{x \to -\infty} F(x) = 0\ and\ \ \lim_{x \to \infty} F(x) = 1$

Types of Continuous distributions

(a) Normal Distribution

If a random variable X is normally distributed, we can write

- $X \sim N(\mu, \sigma^2)$. $Mean = \mu$ and $Variance = \sigma^2$.
- $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Here, $f(x)$ is the pdf
- $F(x) = P(X < x)$. Here, $F(x)$ is the cdf

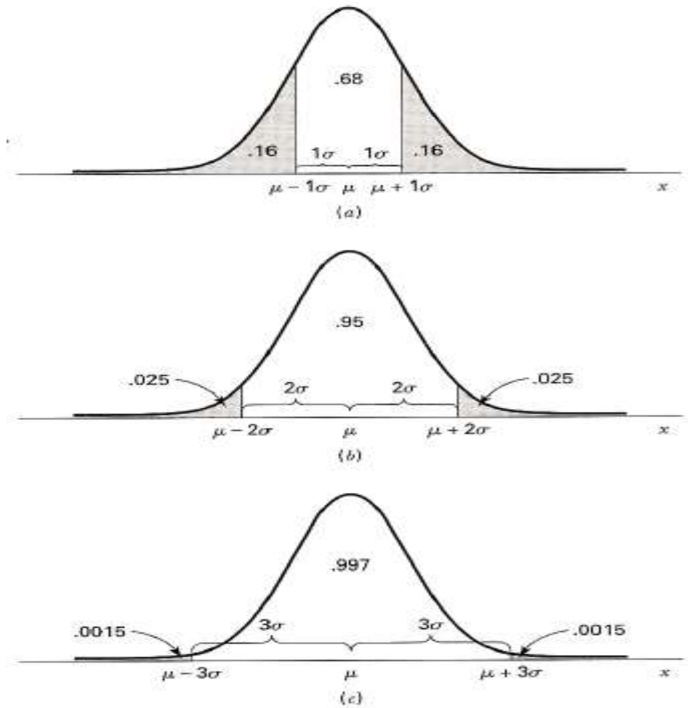


Facts

- Symmetric about its mean μ. That is, $P(X \le \mu) = P(X \ge \mu) = 0.5$ and $P(X \le \mu - a) = P(X \ge \mu - a)$ for any constant a.

- Mean = median = mode

- Area under the curve = 1



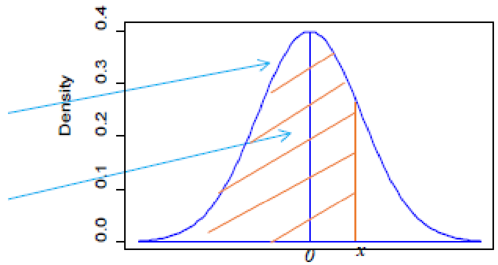
Empirical Formula

68 –95 –99.7 rule (for how much of the distribution is within 1, 2 and 3σ’sfrom the center μ)

(b) Standard normal distribution

This is a special case of Normal distribution. Here, $Mean = \mu = 0$ and $Variance = \sigma^2 = 1$. A standard normal random variable is usually represented by Z. It is also called standardized score or z-score. Thus,

- $Z \sim N(0,1)$.
- $f(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{z^2}{2}}$. Here, $f(z)$ is the pdf
- $\Phi(z) = P(Z < z)$. Here, $\Phi(z)$ is the cdf



For standard normal distribution Table see Page A38 -A39

Stata functions for

- cdf of standard normal distribution: normal(z)

Code:

(a) disp normal(z) for $\Phi(z) = P(Z < z)$
(b) disp 1 – normal(z) for $1 - \Phi(z) = P(Z > z)$
(c) $disp\ normal(z_2)$ – $disp\ normal(z_1)$ for $P(z_1 < Z < z_2)$

Examples. Practice these on Stata.

- P(Z ≤2.45) = Φ(2.45) = disp normal(2.45)
- P(Z >2.45) = 1 - Φ(2.45) = disp 1 - normal(2.45)
- P(0.84 < Z ≤2.45) = Φ(2.45) –Φ(0.84) = disp normal(2.45) – normal(.84)

More on this next Friday ………

- pdf of standard normal distribution: normalden(z)