Random Number Test

SNU/NUKE/EHK

Random Number Generator

- Desirable properties:
  - $E[R] \rightarrow 1/2$, $var[R] \rightarrow 1/12$ as $m \rightarrow \infty$.
- Proof

For a full period LCG, every integer value from 0 to $m-1$ is represented. Thus

$$E = [\{0+1+\cdots+(m-1)\}/m]/m = \{(m-1)(m)/2\}/m^2$$

$$= (m^2-m)/(2m^2) = (1/2)-(1/2m) \rightarrow 1/2 \text{ as } m \rightarrow \infty.$$

$$V = [\{0^2+1^2+2^2+\cdots+(m-1)^2\}/m^2]/m - E^2$$

$$= [(m)(m-1)(2m-1)/6]/m^3 - [(1/2) - (1/2m)]^2$$

$$= [(1/12) - (1/12m^2)] \rightarrow 1/12 \text{ as } m \rightarrow \infty.$$

Uniformity Test

Uniformity Test for Random Numbers

- In testing for uniformity, the hypotheses are as follows:

$$H_0: R_i \sim U[0,1]$$
$$H_1: R_i \nsim U[0,1]$$

- The null hypothesis, H_0, reads that the numbers are distributed uniformly on the interval [0,1].

Uniformity Test for Random Numbers (cont.)

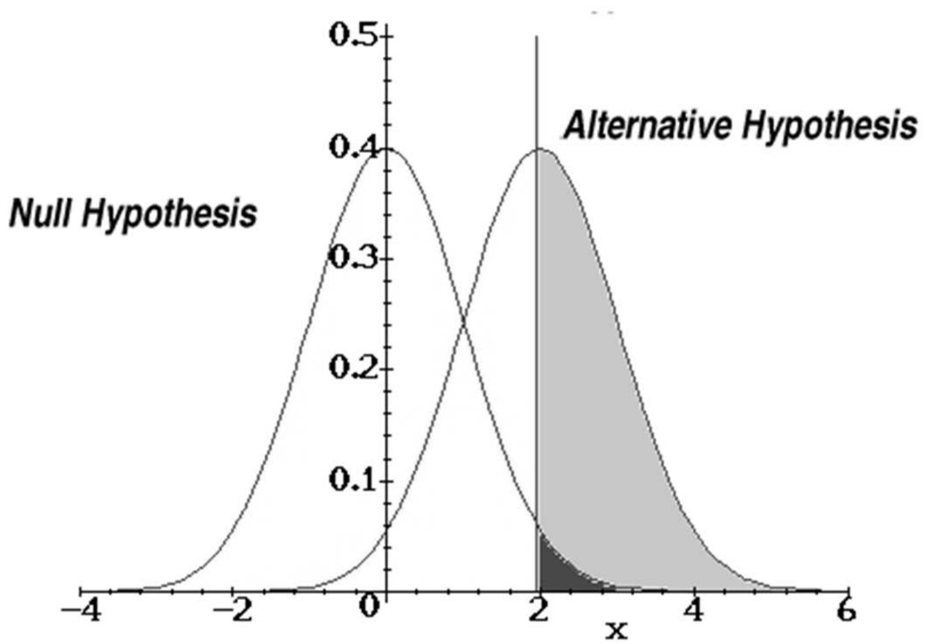
- Significance level, α:

  $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$,

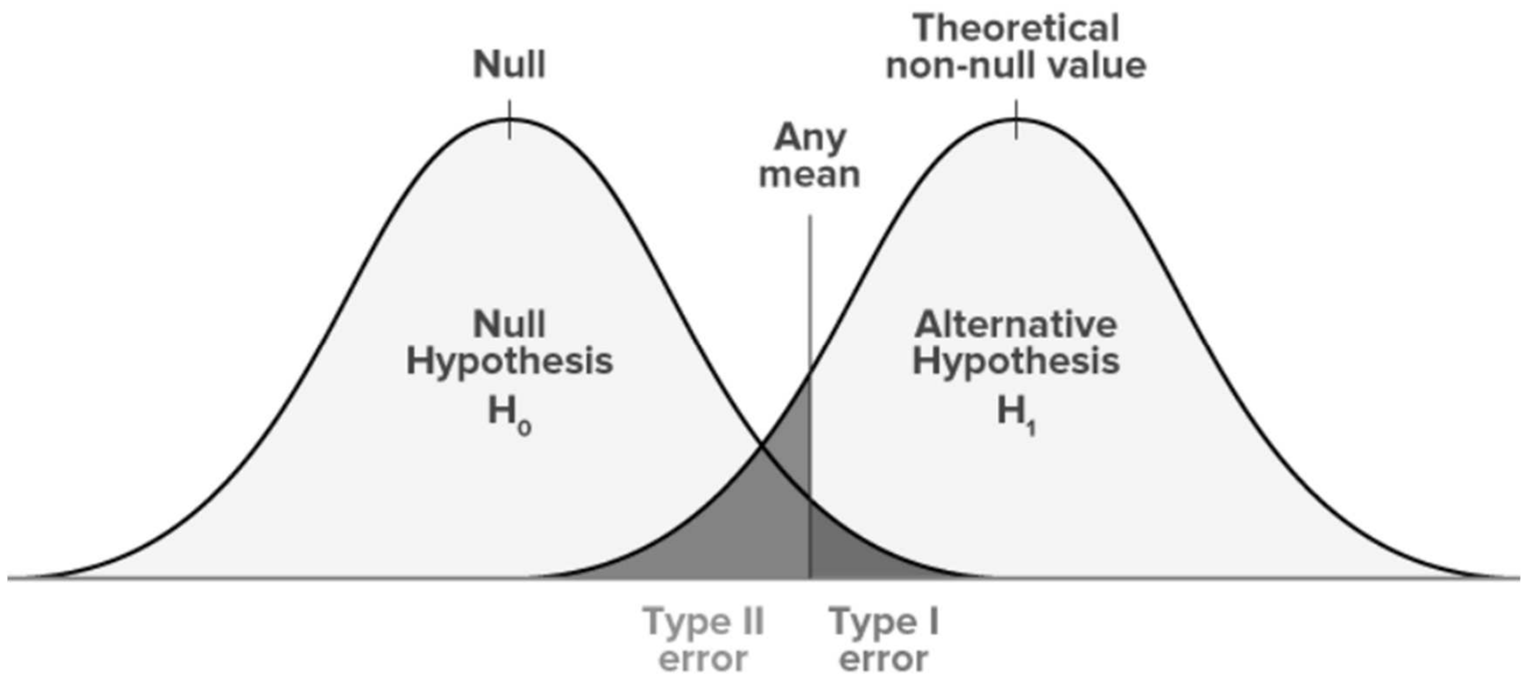
  frequently set to 0.01 or 0.05.

- Confidence interval: $(1-\alpha)100$ %

	Hypothesis	
	Actually True	Actually False
Accept	$1-\alpha$	β (type II error)
Reject	α (type I error)	$1-\beta$



0.5
0.4
0.3
0.2
0.1
Alternative Hypothesis
Null Hypothesis
−4
−2
0
2
4
6
x



Null
Theoretical
non-null value
Any
mean
Null
Hypothesis
H_0
Alternative
Hypothesis
H_1
Type II
error
Type I
error

Uniformity Test

- Frequency (or Spectral) test

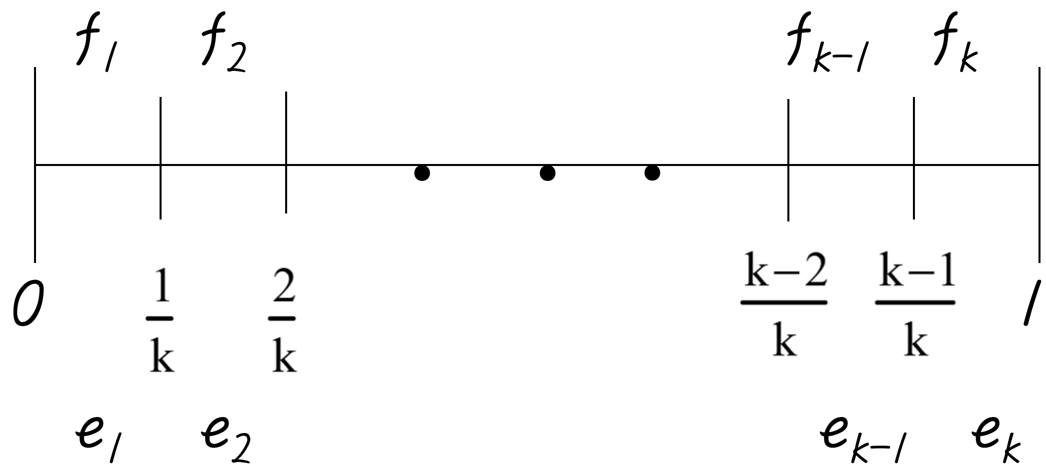
  Uses the Kolmogorov–Smirnov or the chi-squared test to compare the distribution of the set of numbers generated with a uniform distribution.

Uniformity Test for Random Numbers (cont.)

- X^2 Goodness of fit test
  1. Divide n observations into k intervals
  2. Count frequencies f_i, $i=1,2,\cdots,k$ for each interval
  3. Compute

$$\text{not just } V^2 = \sum_{i=1}^{k}(f_i - e_i)^2, \text{ but } X^2 = \sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i}$$

where e_i = expected frequency in the i-th interval and ($/e_i$) is applied to give correct weights to each squared discrepancies.



p_i = expected probabilities observed in interval i

$= e_i/n$ for $i = 1, 2, \cdots, k$

NOTE

- ✓ $(f_i - np_i)/(np_i)^{1/2}$ is the $N(0,1)$ approximation of a multinomial distribution for large n, where

$$E[f_i] = np_i \text{ and } Var\,[f_i] = np_i \cdot (1-p_i).$$

- ✓ For large n, X^2 is approximated to χ^2 distribution with k−1 degrees of freedom

- ✓ Reject randomness on condition $X^2 > \chi^2$

- ✓ $X^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i}$

$$\sim [(k-1)s^2/\sigma^2] \text{ for small identical } p_i\text{'s}$$

$$\text{so that } \sigma^2 \sim \sigma_i^2 = np_i\,(1-p_i) \sim np_i$$

Chi-Squared (Goodness-of-Fit) Test

- *used to test if a sample of data came from a population with a specific distribution.*
- *can be applied to any univariate distribution for which one can calculate the cdf.*
- *can be applied to discrete distributions such as the binomial and the Poisson.*
- *can perform poorly for small sample sizes due to its test statistic not having an approximate chi-squared distribution.*

Chi-Squared Test (cont.)

- Test for the null hypothesis H_0

– H_0 : The data follow a specified distribution f

– H_a : The data do not follow the specified distribution

– Test statistic : $X^2 = \sum_{i=1}^{k}(O_i - E_i)^2/E_i$ $(= \sum_{i=1}^{k} {Z_i}^2)$

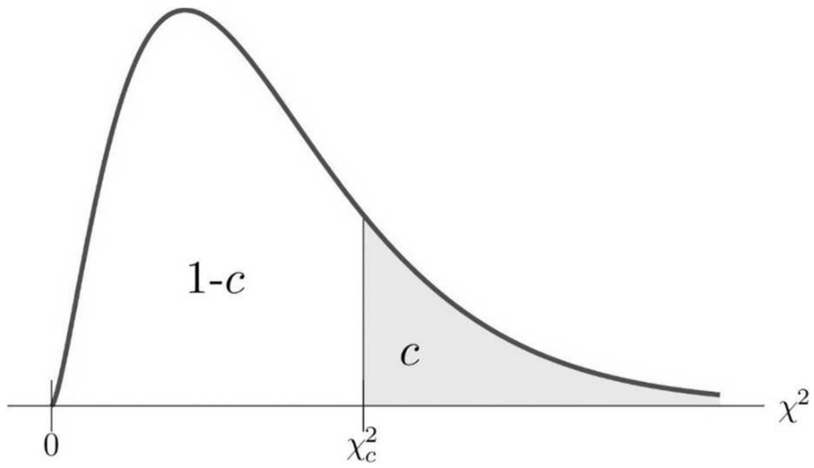
where O_i = the observed frequency for bin i and
E_i = the expected frequency for bin i.

– Significance level : α

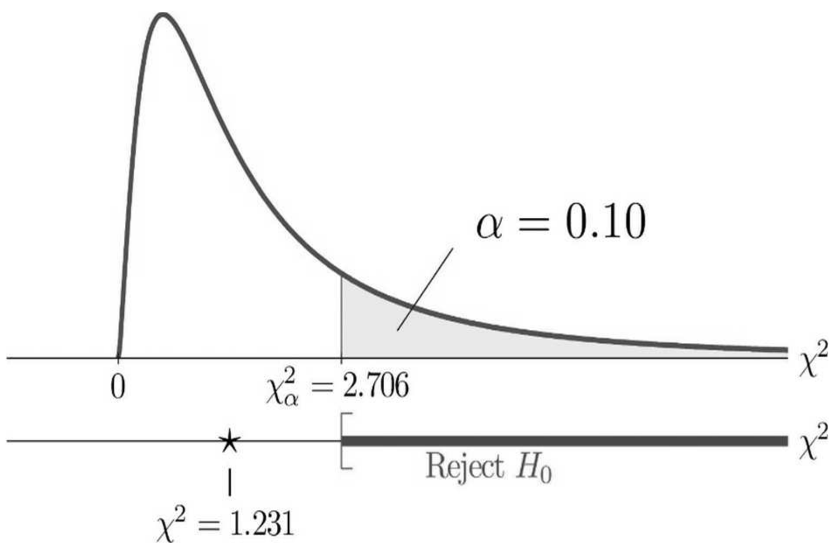
– Critical region : The null hypothesis is rejected if

$$X^2 > {\chi^2}_{\alpha,k-1}$$

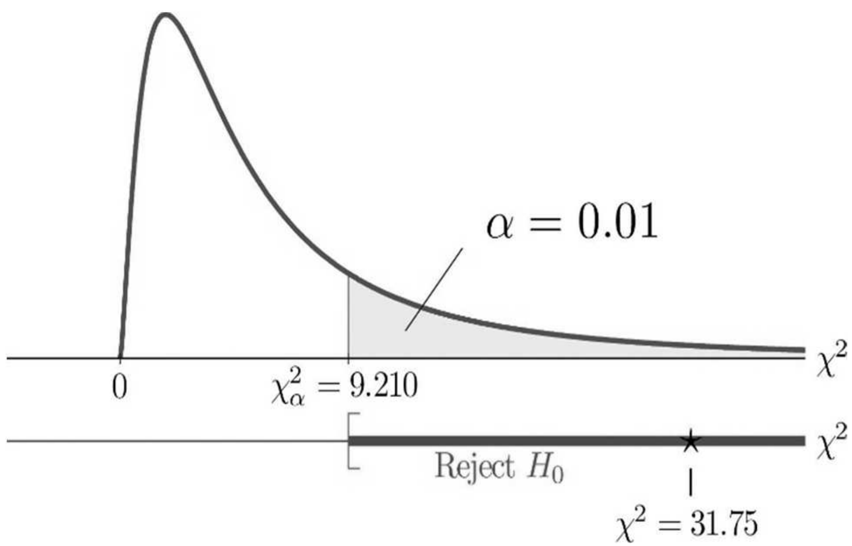
(${\chi^2}_{\alpha,k-1}$ = the 1−α quantile of χ^2 distribution with k−1 degrees of freedom)



1-c
c
0
χ^2_c
χ^2



$\alpha = 0.10$
0
$\chi^2_\alpha = 2.706$
χ^2
Reject H_0
$\chi^2 = 1.231$



$\alpha = 0.01$
0
$\chi^2_\alpha = 9.210$
χ^2
Reject H_0
$\chi^2 = 31.75$

Chi-Squared Test (cont.)

- ✓ The expected frequency is calculated by

$$E_i = N \cdot (F(Y_u) - F(Y_l))$$

where F is the cdf for the distribution f being tested, Y_u is the upper limit for class i, and Y_l is the lower limit, and N is the sample size.

- ✓ The test statistic $\chi^2 = \sum_{i=1}^{k}(O_i - E_i)^2/E_i$ approximately follows a chi-square distribution with (k–c) degrees of freedom, where k is the number of non-empty cells and c is (the number of estimated, not exact, parameters for the distribution + 1).

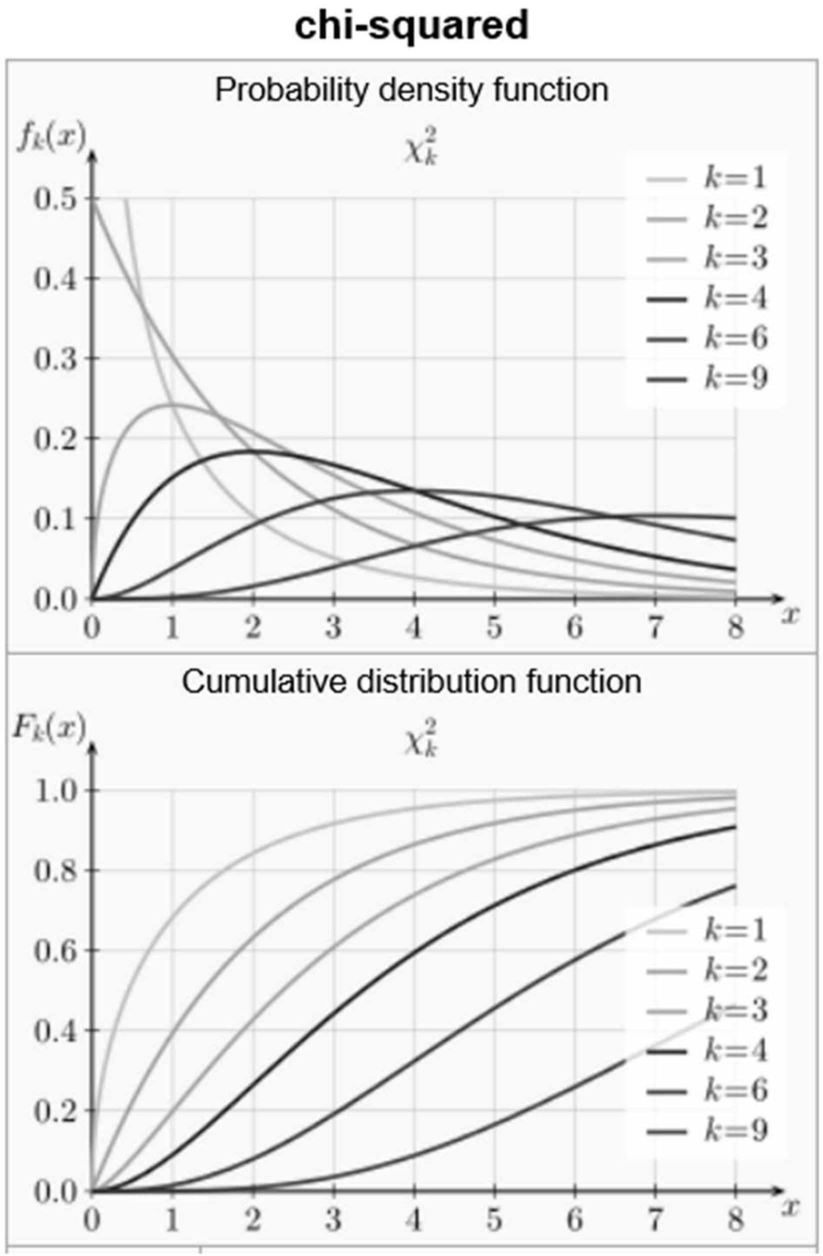
✓Chi-Square Distribution

✓ The χ^2 distribution with k degrees of freedom
= the distribution of a sum of the squares of k independent standard normal random variables.

$$f(x;k) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-x/2} \quad (x \geq 0)$$

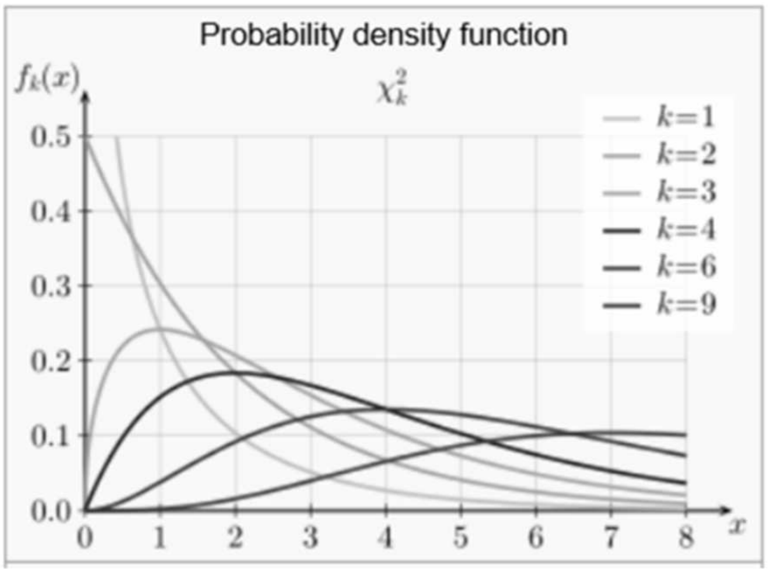
✓ degree of freedom

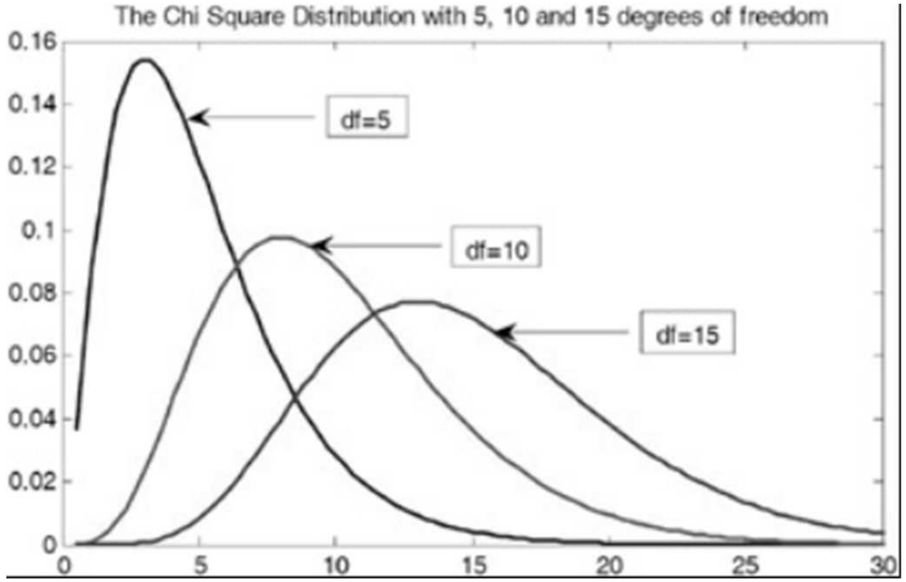
= the number of values in the final calculation of a statistic that are free to vary.



✓Chi-Squared Distribution (cont.)

- ✓ No negative variable values (X^2)
- ✓ Mean (of X^2) is equal to the degrees of freedom
- ✓ As the degree of freedom increases, the standard deviation increases so the chi-square curve spreads out more.
- ✓ As the degree of freedom becomes vary large, the shape becomes more like the normal distribution.





✓Chi-Squared Distribution (cont.)

chi-square distribution table

χ^2_p = the χ^2 value such that the area to its right is p.

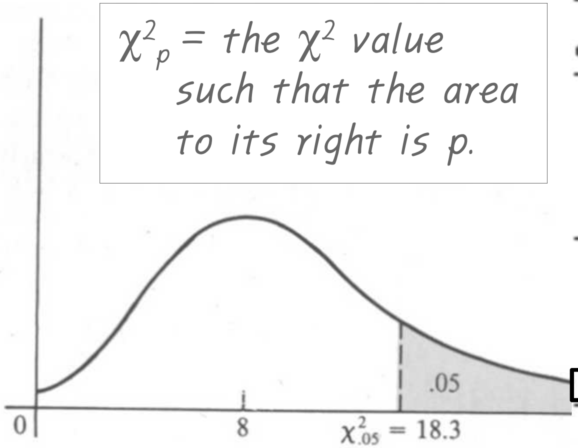


FIGURE 4.5
$P[X^2_{10} \geq \chi^2_{.05}] = .05$ and $P[X^2_{10} < \chi^2_{.05}] = .95$.

df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

- p = the probability that a random sample from a true Poisson distribution would have a larger value of χ^2 than the specified value shown in the table.
  - Very low values (say less than 0.02) indicate abnormally large fluctuations in the data whereas very high probabilities (greater than 0.98) indicate abnormally small fluctuation.
  - Perfect fit to the Poisson distribution for large samples would yield a probability 0.50.

Why not Chi squared test but K-S test

- ✓ Chi square test assumes that the situations produce "normal" data that differ only in that the average outcome in one situation is different from the average outcome in the other situation.
- ✓ If one applies the chi squared test to non-normal data, the risk of error is probably increased.
- ✓ The Central Limit Theorem shows that the chi squared test can avoid becoming unusually fallible when applied to non-normal datasets, if the control/treatment datasets are sufficiently "large".

Kolmogorov-Smirnov (Goodness-of-Fit) Test

- *used as an alternative to the chi-square test when the sample size is small.*
- *A non-parametric and distribution-free test*
- *used to compare a sample with a reference probability distribution (one-sample K-S test) or to compare two samples from the same probability distribution (two-sample K-S test).*
- *does not depend on the underlying cumulative distribution function being tested.*

Kolmogorov–Smirnov Test (cont.)

- Test for the null hypothesis H_0

– H_0 : The data follow a specified distribution f

– H_a : The data do not follow the specified distribution

– Test statistic : $D = Max|F(x_i) - E(x_i)|$

where $F(x_i)$ = the theoretical (exact, not approximate) cdf for the distribution f and $E(x_i)$ = the empirical cdf evaluated, both at x_i.

– Significance level : α

– Critical region : The null hypothesis is rejected if

$$D > CV(\alpha, n) \text{ from K–S distribution}$$

Kolmogorov-Smirnov Test (cont.)

✓ Those two cdf functions evaluated at x_i are defined as

$$F(x_i) = P(X \le x_i)$$

and

$$E(x_i) = \frac{\#\,of\ X's \le x_i}{n} = \frac{i}{n}, \qquad \text{for } I = 1, 2, \cdots, n$$

* $F(x) = x,\ 0 \le x \le 1$ for uniform distribution $f(x)$

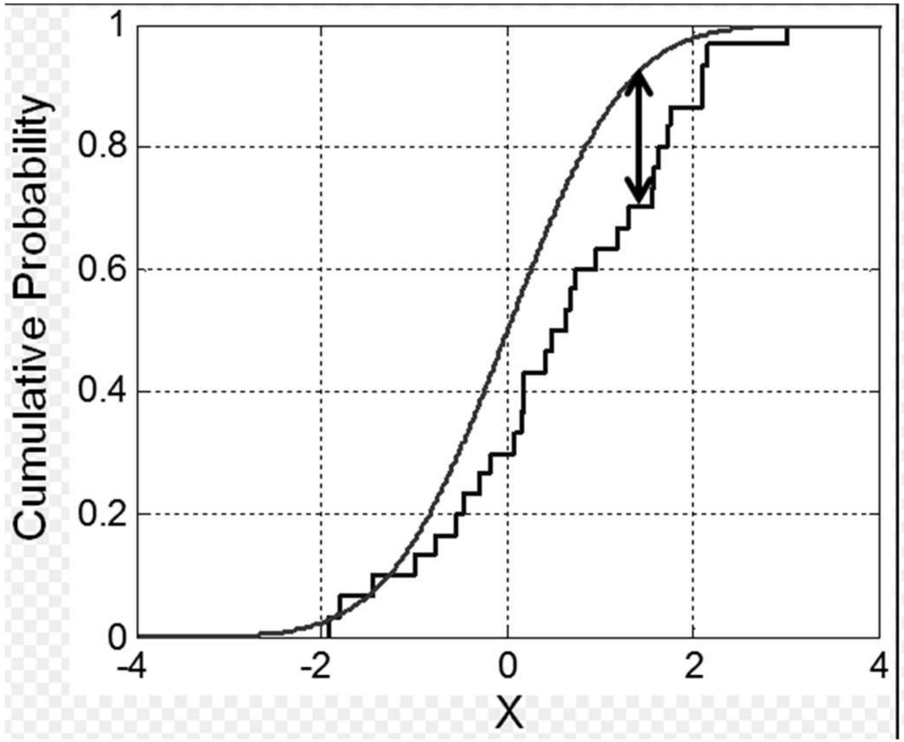
✓ If $D > CV(\alpha, n)$, it is unlikely that $F(x)$ is the underlying data distribution.

✓ The probability of $D > CV(\alpha, n)$ is α.

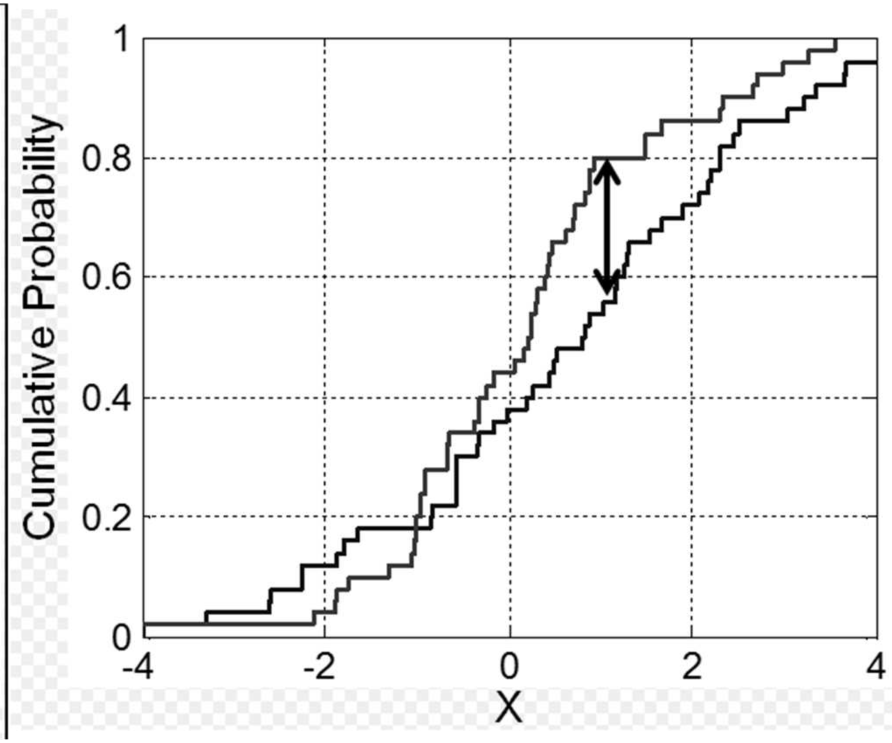
✓ Example of CV table

Table 1 Critical values, $CV(\alpha, n)$, of the KS test with sample size n at the different levels of α.

	Level of significance (α)					
n	0.40	0.20	0.10	0.05	0.04	0.01
5	0.369	0.447	0.509	0.562	0.580	0.667
10	0.268	0.322	0.368	0.409	0.422	0.487
20	0.192	0.232	0.264	0.294	0.304	0.352
30	0.158	0.190	0.217	0.242	0.250	0.290
50	0.123	0.149	0.169	0.189	0.194	0.225
>50	$\frac{0.87}{\sqrt{n}}$	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.37}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$



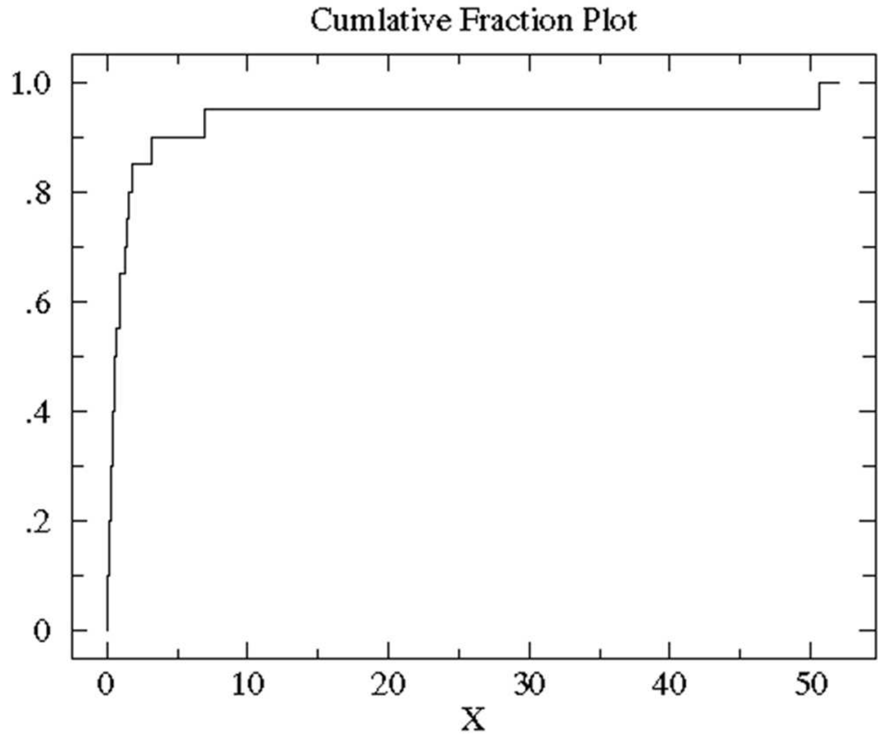
One-sample Kolmogorov–Smirnov statistic: Red line is CDF; blue line is an ECDF (empirical CDF); and the black arrow is the K-S statistic.

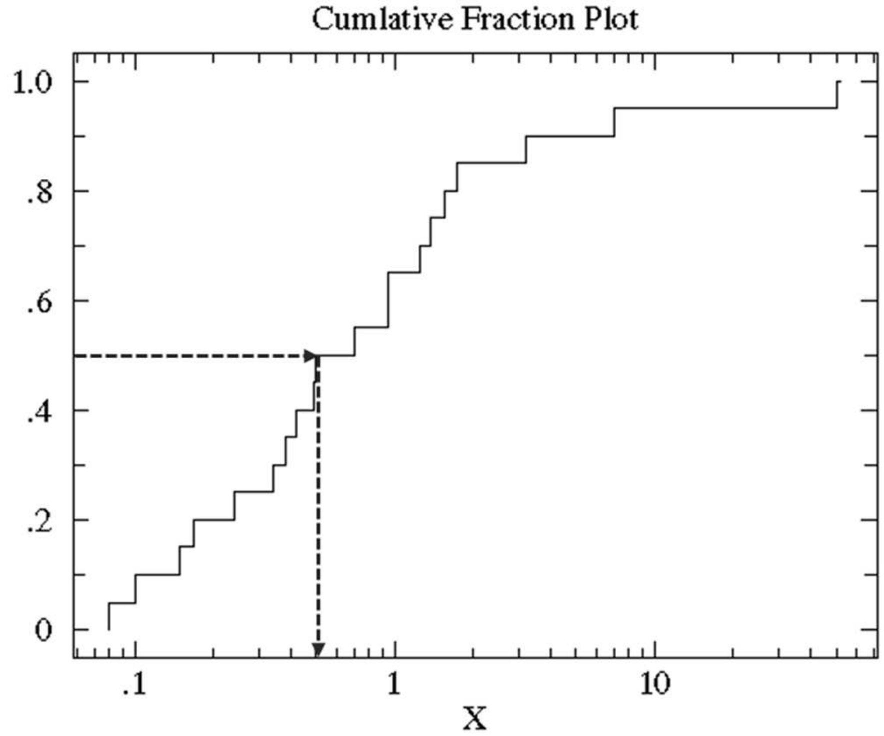


two-sample Kolmogorov–Smirnov statistic: Red and blue lines each correspond to empirical distribution functions, and the black arrow is the two-sample KS statistic.

* Kolmogorov published the asymptotic K-S distribution and K-S statistic and Smirnov published the table of K-S cdf.

Data scale in linear vs. in log

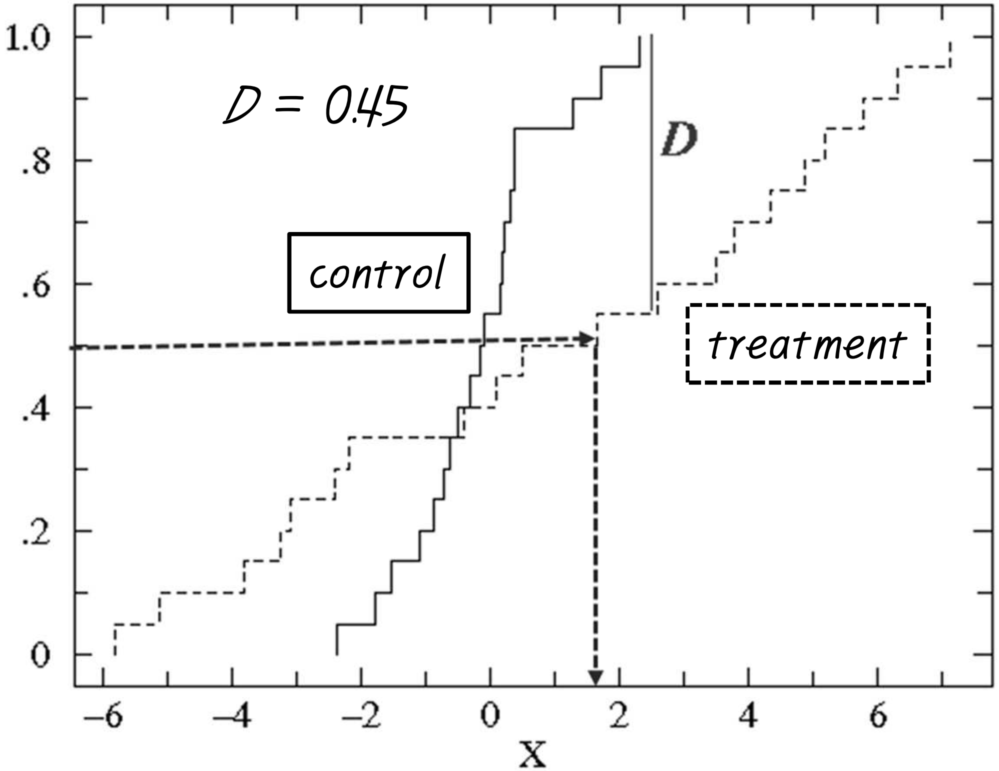




Cumulative Fraction Plot: example #1

Control A={0.22, −0.87, −2.39, −1.79, 0.37, −1.54, 1.28, −0.31, −0.74, 1.72, 0.38, −0.17, −0.62, −1.10, 0.30, 0.15, $\boxed{2.30}$, 0.19, −0.50, −0.09}

Treatment A={−5.13, −2.19, −2.43, −3.83, 0.50, −3.25, 4.32, 1.63, 5.18, −0.43, $\boxed{7.11}$, 4.87, −3.10, −5.81, 3.76, 6.31, 2.58, 0.07, 5.76, 3.50}

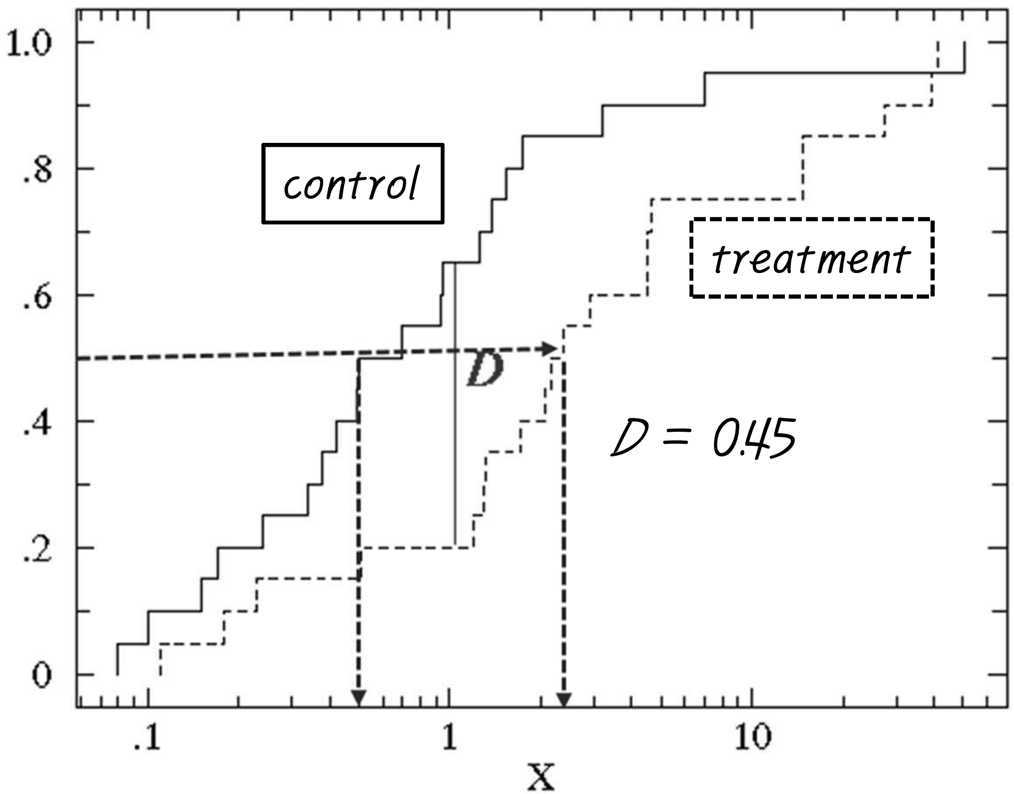


- ✓ Both data sets do not differ in mean, but differ in variance.
- ✓ Chi square test does not see the difference.

Cumulative Fraction Plot: example #2

Control B = {1.26, 0.34, 0.70, 1.75, 50.57, 1.55, 0.08, 0.42, 0.50, 3.20, 0.15, 0.49, 0.95, 0.24, 1.37, 0.17, 6.98, 0.10, 0.94, 0.38}

Treatment B= {2.37, 2.16, 14.82, 1.73, 41.04, 0.23, 1.32, 2.91, 39.41, 0.11, 27.44, 4.51, 0.51, 4.50, 0.18, 14.68, 4.66, 1.30, 2.06, 1.19}



- ✓ Both data sets were drawn from lognormal (that is, non-normal) distributions that differ in median.
- ✓ One can "see" the difference in the plot.

Cumulative Fraction vs. Percentile

✓ Take a data set

{−0.45, 1.11, 0.48, −0.82, −1.26}

✓ Sort from the smallest to the largest:

{ −1.26, −0.82, −0.45, 0.48, 1.11 }

✓ Calculate the percentiles:

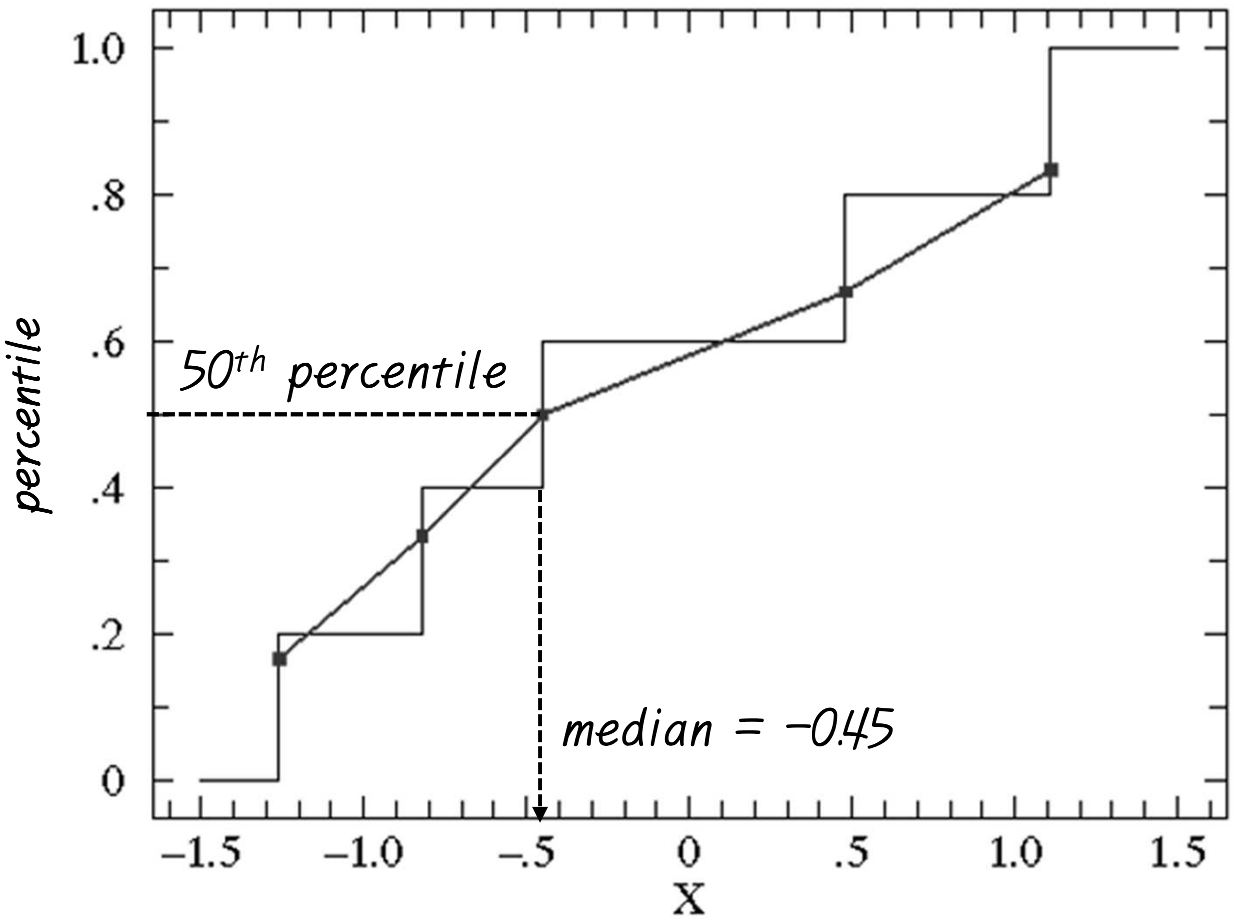
Percentile = $r/(N+1)$ X 100 (- th)

where r is the location of each point among N data

✓ Align the set of (datum, percentile) pairs

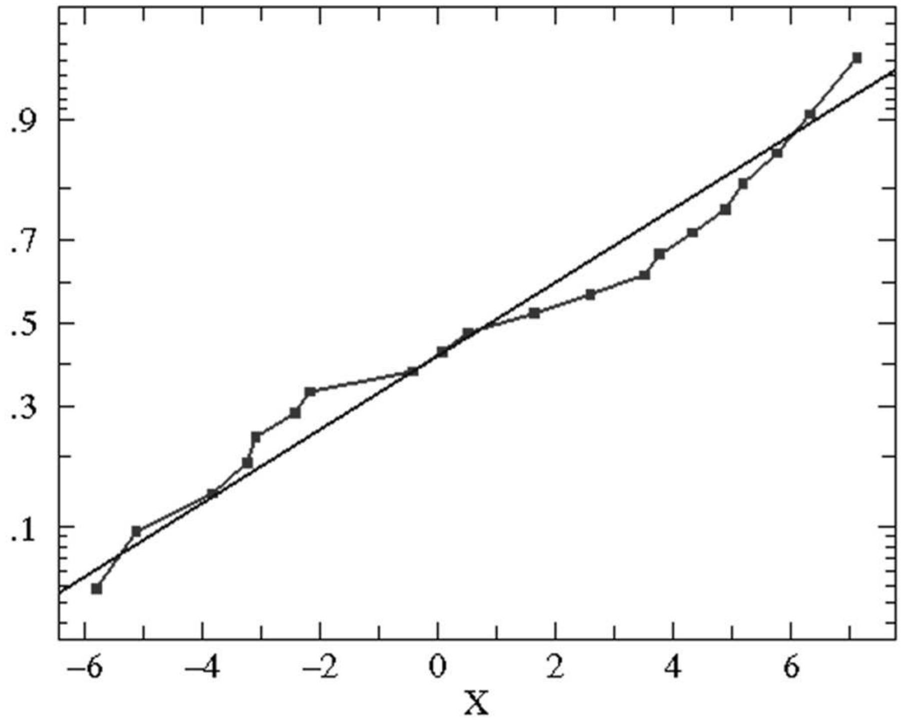
{ (−1.26,.167), (−0.82,.333), (−0.45,.5), (0.48,.667), (1.11,.833) }

percentile plot vs. cumulative fraction

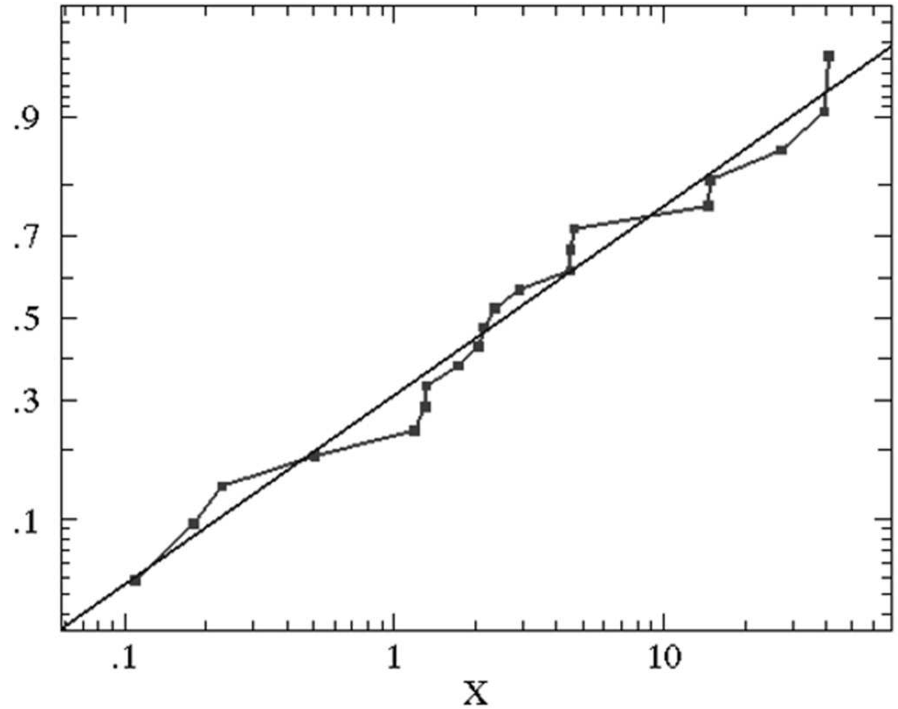


percentile plot on probability graph paper

linear-probability scale



log-probability scale

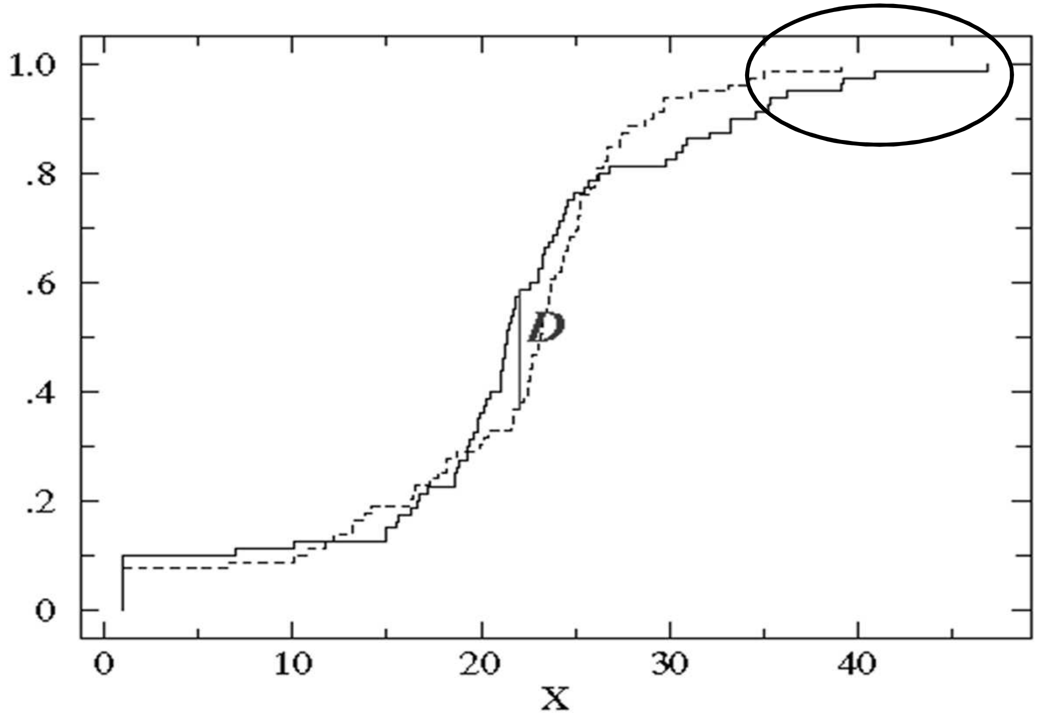


- ✓ Uniformly distributed data will plot as a straight line using the regular graph paper.
- ✓ Normally-distributed data will plot as a straight line on the linear-probability paper.
- ✓ Lognormal data will plot as a straight line with log-probability scaled axes.

Chi Square test vs. Kolmogorov-Smirnov test

Control C={23.4, 30.9, 18.8, 23.0, 21.4, 1, 24.6, 23.8, 24.1, 18.7, 16.3, 20.3, 14.9, 35.4, 21.6, 21.2, 21.0, 15.0, 15.6, 24.0, 34.6, 40.9, 30.7, 24.5, 16.6, 1, 21.7, 1, 23.6, 1, 25.7, 19.3, 46.9, 23.3, 21.8, 33.3, 24.9, 24.4, 1, 19.8, 17.2, 21.5, 25.5, 23.3, 18.6, 22.0, 29.8, 33.3, 1, 21.3, 18.6, 26.8, 19.4, 21.1, 21.2, 20.5, 19.8, 26.3, 39.3, 21.4, 22.6, 1, 35.3, 7.0, 19.3, 21.3, 10.1, 20.2, 1, 36.2, 16.7, 21.1, 39.1, 19.9, 32.1, 23.1, 21.8, 30.4, 19.62, 15.5}

Treatment C={16.5, 1, 22.6, 25.3, 23.7, 1, 23.3, 23.9, 16.2, 23.0, 21.6, 10.8, 12.2, 23.6, 10.1, 24.4, 16.4, 11.7, 17.7, 34.3, 24.3, 18.7, 27.5, 25.8, 22.5, 14.2, 21.7, 1, 31.2, 13.8, 29.7, 23.1, 26.1, 25.1, 23.4, 21.7, 24.4, 13.2, 22.1, 26.7, 22.7, 1, 18.2, 28.7, 29.1, 27.4, 22.3, 13.2, 22.5, 25.0, 1, 6.6, 23.7, 23.5, 17.3, 24.6, 27.8, 29.7, 25.3, 19.9, 18.2, 26.2, 20.4, 23.3, 26.7, 26.0, 1, 25.1, 33.1, 35.0, 25.3, 23.6, 23.2, 20.2, 24.7, 22.6, 39.1, 26.5, 22.7}



- ✓ The Chi square test can not see the difference (large N), whereas the KS-test can.
- ✓ Take the Cauchy distribution instead of Normal distribution!